

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1236	100.0	1239	6	BD274350	BD274350 Candida a
2	1236	100.0	2060	6	BD274349	BD274349 Candida a
3	1232.8	99.7	1239	6	AX489193	AX489193 Sequence
4	217.4	17.6	1560	8	YSCNOPEAT	M90638 Saccharomyc
5	217.4	17.6	1739	8	YSCFTFIIIA	M80611 Saccharomyc
6	217.4	17.6	37497	8	YSCPE977	U25841 Saccharomyc
7	199.2	16.1	2133	8	YSCRP026A	M33924 S.cerevisia
8	193	15.6	797	11	CNS0801NG	AL400514 T7 end of
9	101	8.2	169063	9	AC138126	AC138126 Homo sapi
10	101	8.2	177239	9	AC073544	AC073544 Homo sapi
11	99.4	8.0	2132	6	AX714042	AX714042 Sequence
12	99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
13	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
14	92.4	8.0	169500	2	AC138469	AC138469 Homo sapi
15	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
16	97.8	7.9	1203	9	D70831	D70831 Homo sapien
17	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
18	97.2	7.9	115995	9	AC011494	AC011494 Homo sapi
19	96.6	7.8	2320	6	AR270491	AR270491 Sequence
20	96.6	7.8	2320	9	HSU35376	U35376 Human repre
21	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
22	95	7.7	381	9	HSZFP9	X78932 H.sapiens H
23	94.6	7.7	186233	9	AC092329	AC092329 Homo sapi
24	94.6	7.7	189317	2	AC024483	AC024483 Homo sapi
25	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
26	93.6	7.6	1389	9	M27879	M27879 Homo sapien
27	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
28	93.4	7.6	199585	10	AC124426	AC124426 Mus muscu
29	92.4	7.5	41153	9	AC016528	AC016528 Homo sapi
30	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
31	92	7.4	2873	9	HUMHPLK	M55422 Human Kruep
32	92	7.4	2873	11	HWMSWS1269	G28705 SMS3376 Er
33	92	7.4	2873	11	HWMSWS1269	G18281 human chrom
34	92	7.4	156835	9	AC010620	AC010620 Homo sapi
35	92	7.4	191082	2	AC141066	AC141066 Homo sapi
36	92	7.4	203396	9	AC073310	AC073310 Homo sapi
37	91.8	7.4	1870	9	AY044432	AY044432 Homo sapi
38	91.8	7.4	2110	6	AX748253	AX748253 Sequence
39	91.8	7.4	2110	9	AK093669	AK093669 Homo sapi
40	91.8	7.4	2909	9	HSM808021	HSM808021 Homo sapi
41	91.8	7.4	2976	9	BC036439	BC036439 Homo sapi
42	91.8	7.4	3078	6	AX747097	AX747097 Sequence
43	91.8	7.4	3078	9	AK091618	AK091618 Homo sapi
44	91.8	7.4	3123	9	BC040594	BC040594 Homo sapi
45	91.8	7.4	174994	9	AC022145	AC022145 Homo sapi

ALIGNMENTS

RESULT 1	BD274350	1239 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD274350				
DEFINITION	Candida albicans tflIIAgene (CatfIIIA) and the coded CATfIIIA protein.				
ACCESSION	BD274350				
VERSION	BD274350.1				
KEYWORDS	JP 2002531068-A/2.				
SOURCE	Candida albicans				
ORGANISM	Candida albicans				
REFERENCE	1 (bases 1 to 1239)				
AUTHORS	Saccharomycetales; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
TITLE	Pallier F.B., Camier S. and Sentenac A.				
	Candida albicans tflIIAgene (CatfIIIA) and the coded CATfIIIA				

Pred. No. is the number of results predicted by chance to have a

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: May 9, 2004, 03:49:53 ; Search time 5075.84 Seconds  
(without alignments)  
10554.310 Million cell updates/sec  
Title: US-09-831-804-1\_COPY\_720\_1955  
Perfect score: 1236  
Sequence: 1 atagtgaaagtgcgaac.....aaacatcagtgattttctga 1236  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.cm.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sv.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Search Nucleo



```
CC      Key      protein      Location/Qualifiers
FH      source      1..2060
FT      /organism="Candida albicans"
FT      /organism="Candida albicans"
FEATURES
    source
        Location/Qualifiers
        1..2060
        /organism="Candida albicans"
        /mol_type="Genomic DNA"
        /db_xref="taxon:5476"
ORIGIN
    Query Match      100.0%; Score 1236; DB 6; Length 2060;
    Best Local Similarity 100.0%; Pred. No. 3.8e-190;
    Matches 1236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATGAGTGAAGTGAAGAAACCAATCGATATCATCTTTTAATATCTTTCTTTCTTCATCA 60
DB      720 ATGAGTGAAGTGAAGAAACCAATCGATATCATCTTTTAATATCTTTCTTTCTTCATCA 779
QY      61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTATAAAGCCATATAATCGACCATCA 120
DB      780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTATAAAGCCATATAATCGACCATCA 839
QY      121 TTATTAGACCAACATTAAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
DB      840 TTATTAGACCAACATTAAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899
QY      181 GATTGTGAATACATTTTTCAGAAATCACATTTTGGAAACACATATTTCTATCATATCC 240
DB      900 GATTGTGAATACATTTTTCAGAAATCACATTTTGGAAACACATATTTCTATCATATCC 959
QY      241 GAAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTCGACCAACATTTG 300
DB      960 GAAAAAAAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTCGACCAACATTTG 1019
QY      301 AAAAGACATGAATCCACCCATCAAAAGTCTTTAAATGTACATTTGAAATTTGTCAAGAA 360
DB      1020 AAAAGACATGAATCCACCCATCAAAAGTCTTTAAATGTACATTTGAAATTTGTCAAGAA 1079
QY      361 GCATTTTAAACATCAATCTTTAAAGACATATATATTTCTGTCATGAAATTTGAAATTTG 420
DB      1080 GCATTTTAAACATCAATCTTTAAAGACATATATATTTCTGTCATGAAATTTGAAATTTG 1139
QY      421 ACCTGTAAACATGTATTAAGATTTTTCATCGACCTTCAAAATTTAGACCAACATTAATTA 480
DB      1140 ACCTGTAAACATGTATTAAGATTTTTCATCGACCTTCAAAATTTAGACCAACATTAATTA 1199
QY      481 AAACATCATGTCGATCTCTCTTATCAATGTGATCATCTCTGTTGTTTAAATTTTC 540
DB      1200 AAACATCATGTCGATCTCTCTTATCAATGTGATCATCTCTGTTGTTTAAATTTTC 1259
QY      541 CAAACITGGTCAGTATTACAAATTTTCATATAAAACAACTGCATCCAAAACCTTAAATGCTCT 600
DB      1260 CAAACITGGTCAGTATTACAAATTTTCATATAAAACAACTGCATCCAAAACCTTAAATGCTCT 1319
QY      601 AATGTGGTAAAGTGTGTGGGAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 560
DB      1320 AATGTGGTAAAGTGTGTGGGAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379
QY      661 GATTCACCATGATCAAAATATGACCTTGTGATATTTGATCTGGGAAAATTTGCAAG 720
DB      1380 GATTCACCATGATCAAAATATGACCTTGTGATATTTGATCTGGGAAAATTTGCAAG 1439
QY      721 AAAAATGAATAGTTGAACATATATAATCTTCATGATGGTAAATATCCCTGATGATTTA 780
DB      1440 AAAAATGAATAGTTGAACATATATAATCTTCATGATGGTAAATATCCCTGATGATTTA 1499
QY      781 TTAAGGAACTGAAGTGAAATATTAGAGAACTTATAGATCAAGGATCGAAATTTAAAT 840
DB      1500 TTAAGGAACTGAAGTGAAATATTAGAGAACTTATAGATCAAGGATCGAAATTTAAAT 1559
QY      841 AATTTCATGAATTTAGAAACAGAGAAATTAAAGTGGAGAGATGAAGAGATGAAGAA 900
DB      841 AATTTCATGAATTTAGAAACAGAGAAATTAAAGTGGAGAGATGAAGAGATGAAGAA 900
```

```
Db      1560 AATTTCATGAATTTAGAAACAGAGAAATTAAGTGGAGAGATGAAGAGATGAAGAA 1619
QY      901 GATAGTCTAGATGAAAAAGAGTATGTTAGATCAGACTCAATGTGAGTCAAGATCA 960
DB      1620 GATAGTCTAGATGAAAAAGAGTATGTTAGATCAGACTCAATGTGAGTCAAGATCA 1679
QY      961 ATAAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTCTTAAACTTATTTCTGAAT 1020
DB      1680 ATAAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTCTTAAACTTATTTCTGAAT 1739
QY      1021 AGTGGGAAGAGATCAATTTGCTTAAGATAATTTGTGATAGATGTTTCTTAGAGATAT 1080
DB      1740 AGTGGGAAGAGATCAATTTGCTTAAGATAATTTGTGATAGATGTTTCTTAGAGATAT 1799
QY      1081 GATTTAGCTCGACATTTTGAATGCGATGATATAATTTTACAAAGAAATGAGTCAATCTTA 1140
DB      1800 GATTTAGCTCGACATTTTGAATGCGATGATATAATTTTACAAAGAAATGAGTCAATCTTA 1859
QY      1141 AATAGTATAGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGGTTAAAAAGCCAGATG 1200
DB      1860 AATAGTATAGAAAAAGAAAGAACTCCAGAGGTGAACCATTTGGTTAAAAAGCCAGATG 1919
QY      1201 GATTTATTTGCCAAATGAACATCAGTCAATCTCGA 1236
DB      1920 GATTTATTTGCCAAATGAACATCAGTCAATCTCGA 1955
RESULT 3
AX489193      1239 bp      DNA      linear      PAT 16-AUG-2002
LOCUS      Sequence 6493 from Patent WO02053728.
ACCESSION      AX489193
VERSION      AX489193.1 GI:22323205
KEYWORDS
SOURCE      Candida albicans
ORGANISM      Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1 Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
Gene disruption methodologies for drug target discovery
Patent: WO 02053728-A 6493 11-JUL-2002;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
    source
        Location/Qualifiers
        1..1239
        /organism="Candida albicans"
        /mol_type="unassigned DNA"
        /db_xref="taxon:5476"
ORIGIN
    Query Match      99.7%; Score 1232.8; DB 6; Length 1239;
    Best Local Similarity 99.8%; Pred. No. 1.4e-189;
    Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 ATGAGTGAAGTGAAGAAACCAATCGATATCATCTTTTAATATCTTTCTTTCTTCATCA 60
DB      1 ATGAGTGAAGTGAAGAAACCAATCGATATCATCTTTTAATATCTTTCTTTCTTCATCA 60
QY      61 CGTCCCAAAAAGTATTTTGCACATATGAAGGGTGTATAAAGCCATATAATCGACCATCA 120
DB      61 CGTCCCAAAAAGTATTTTGCACATATGAAGGGTGTATAAAGCCATATAATCGACCATCA 120
QY      121 TTATTAGAGCAACATTTAAGAAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
DB      121 TTATTAGAGCAACATTTAAGAAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180
QY      181 GATTGTGATAAAGCATTTTTCAGAAAATCAGATTTGGAACACATATTTGATATCATTTCC 240
DB      181 GATTGTGATAAAGCATTTTTCAGAAAATCAGATTTGGAACACATATTTGATATCATTTCC 240
QY      241 GAAAAAAAACCATTCATTTGTTGAGTGTGTGTAAGGGGTAAATTTCTCGACCAACATTTG 300
DB      241 GAAAAAAAACCATTCATTTGTTGAGTGTGTGTAAGGGGTAAATTTCTCGACCAACATTTG 300
```

KEYWORDS RNA polymerase; transcription factor IIIA.  
SOURCE Saccharomyces cerevisiae (baker's yeast)  
ORGANISM Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 1560)  
AUTHORS Woychik,N.A. and Young,R.A.  
TITLE Genes encoding transcription factor IIIA and the RNA polymerase  
common subunit RPB6 are divergently transcribed in Saccharomyces  
cerevisiae  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)  
MEDLINE 92237295  
PUBMED 1570325  
COMMENT Original source text: Saccharomyces cerevisiae (strain S288C) DNA.  
FEATURES  
Location/Qualifiers  
1..1560  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"  
/strain="S288C"  
/db\_xref="taxon:4932"  
ORIGIN chromosome 14.  
Query Match 17.6%; Score 217.4; DB 8; Length 1560;  
Best Local Similarity 59.4%; Pred. No. 9.9e-26;  
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;  
QY 62 GTCCCAAAAGTATATTTCACATATGAAGGGTGTGATAAGCCTATAATCGACCATCAT 121  
DB 368 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTTCAAGACCTTCAA 427  
QY 122 TATTAGAGCAACATTTTAAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181  
DB 428 TTTTGACTGNACCAATTAAGCGTACATCAGGGTTTA--AGAGCATTTTCAGTGTGATA 484  
QY 182 ATTGTGATAAAGCATTTTTCAGAAAATTCACATTTTGGAAAACACATATGTATCATCTCCG 241  
DB 485 AGTGTGCAAAATCCTTCGTTTAAAGAGTCACCTTAGAGAGACACTTGTATACGCATTTCTG 544  
QY 242 AAAAAAACCATTCCTAGTGTGTGTAAGAGGGTTAATCTTCGACAAACACTTGA 301  
DB 545 ATAGCAAAACCTTCCTCAATGTTCTTATTTGGAAAAGAGTGACGACTCGCCAGCAACTGA 604  
QY 302 AAAGACATGAATTCACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAAG 361  
DB 605 AGGCACACGAAGTAAGCATACCAATCTTTTCAATTTTGTCCAGAGGAAGGATGCAACCTCC 664  
QY 362 CATTTTAAACATCAATCTTTAAGACATCATATATTCTGTTCATGAAAAACATTA 421  
DB 665 GATTCTACAAGCATCCCAATTAAGGCAATTTTATCTGTTCAATTAACATAA 724  
QY 422 CGTGTAAACAATGTATAAAGTTTTTCATCGACCTTCAAAATTAGCAACAATAAATTAA 481  
DB 725 CCTGTCCACACTGCAATAAAGGCTTTCAGAGGCGCATATAGGCTTAAGAAATTCACATTTCTA 784  
QY 482 AACATCA-----TGGTGGATCTCCTCTTATCAATGTGATCATCTCTGGTTGTTTAAAA 535  
DB 785 AACATCAAGTCTCTGAGGTAGAAAATCTCTTACCAGTACTTTTGTGTTGTGCAAG 844  
QY 536 ATTTCCAAAATTGGTCAAGTATTACAAATTCATATAAACAACATGCAATCCAAAATTAAT 595  
DB 845 AGTTTCGGATATGGTCACATTCGCAATTCGCAATATAAATAATGATCATCTCTAAATTAAT 904  
QY 596 GTCTTAAATGTGTAAGGTTGTGTGGGAAAAAGGTTTATCTTCACATATGTTAAGTC 655  
DB 905 GTCTTATTTGTAGCAAAACCATGTGTGGGGAAAATGGTTTACAAATGCACATGATTATTC 964  
QY 656 ATGATGATTTCTACCATCATCAAAATATGGACTTGTGATTAATTTGATGTGGGAAATTTG 715  
DB 965 ATGACGACTACTAGTAACCAAAAATTTGGAAGTGTGCATATATGTCCTGATATGTCITTTT 1024  
QY 716 CAAAGAAAATGAATTAAGTTGAACATTTAT 744  
DB 1025 CTAGAAAACATGATCTTCTCAGCGATTAT 1053

301 AAAAGACATGAATCACCCTACCAAGTCATTTAAATGTACATTTGAAATTTGCAAGAA 360  
DB 301 AAAAGACATGAATCACCCTACCAAGTCATTTAAATGTACATTTGAAATTTGCAAGAA 360  
QY 361 GCATTTTATAACATCAATCTTTAAGACATCATATATTATCTGTTTCATGAAAAAACATTA 420  
DB 361 GCATTTTATAACATCAATCTTTAAGACATCATATATTATCTGTTTCATGAAAAAACATTA 420  
QY 421 ACCTGTAACAATGTAATAAAGTTTTCCTCGACCTTCAAAATTAGCAACAATAATTA 480  
DB 421 ACCTGTAACAATGTAATAAAGTTTTCCTCGACCTTCAAAATTAGCAACAATAATTA 480  
QY 481 AAACATCATGTGGTATCTCTCTTATCAATGTGATCATCTCTGGTGTGTTTAAAAATTTTC 540  
DB 481 AAACATCATGTGGTATCTCTCTTATCAATGTGATCATCTCTGGTGTGTTTAAAAATTTTC 540  
QY 541 CAAACTTGTGATGATTAACAATTCATATAAACAACCTGCATCCAAAATTTAAATGTCCT 600  
DB 541 CAAACTTGTGATGATTAACAATTCATATAAACAACCTGCATCCAAAATTTAAATGTCCT 600  
QY 601 AAATGTGTAAGAGTTGTGTTGGGAAAAAGGTTTATCTTCACATATGTTTAAGTCATGAT 660  
DB 601 AAATGTGTAAGAGTTGTGTTGGGAAAAAGGTTTATCTTCACATATGTTTAAGTCATGAT 660  
QY 661 GATTTCTACCATGATCAAAATATGACCTTGTGATTTTGTGATGTGGGAAAATTTGCAAG 720  
DB 661 GATTTCTACCATGATCAAAATATGACCTTGTGATTTTGTGATGTGGGAAAATTTGCAAG 720  
QY 721 AAAATGATTAAGTGAACATTAATATCTTCCTGATGTTGTAATCCCTGATGATTTA 780  
DB 721 AAAATGATTAAGTGAACATTAATATCTTCCTGATGTTGTAATCCCTGATGATTTA 780  
QY 781 TTAAGGAACTGAAGTGAAGAAATTTAGAGAACTTATTAGATCAAGGATCGAAATTTAAAT 840  
DB 781 TTAAGGAACTGAAGTGAAGAAATTTAGAGAACTTATTAGATCAAGGATCGAAATTTAAAT 840  
QY 841 AATTTGCATGAATTAGAAACAGAGAAATTTAAAGTGGAGAGATGAAGAGATGAAGAA 900  
DB 841 AATTTGCATGAATTAGAAACAGAGAAATTTAAAGTGGAGAGATGAAGAGATGAAGAA 900  
QY 901 GATAGTCTAGATGAAAAAGAGTGTGTTAGATCAGACTCAATGTGAGCTCAAAAGATCA 960  
DB 901 GATAGTCTAGATGAAAAAGAGTGTGTTAGATCAGACTCAATGTGAGCTCAAAAGATCA 960  
QY 961 ATAAATCATTTACTGCTTCTTTGGAGGTTTCAAGAGTGTTCCTAACTTATTTCTGAAT 1020  
DB 961 ATAAATCATTTACTGCTTCTTTGGAGGTTTCAAGAGTGTTCCTAACTTATTTCTGAAT 1020  
QY 1021 AGTGGGAAGAGATCAATTTGCTTGAAGATAATTTGATGAGATGTTTCTAGAGAAAT 1080  
DB 1021 AGTGGGAAGAGATCAATTTGCTTGAAGATAATTTGATGAGATGTTTCTAGAGAAAT 1080  
QY 1081 GATTTACGTCGACATTTGAAATGGCATGATGATTAATTTCAAGAAATTTGAGTCATTTA 1140  
DB 1081 GATTTACGTCGACATTTGAAATGGCATGATGATTAATTTCAAGAAATTTGAGTCATTTA 1140  
QY 1141 AATAGTATGAAAAAGAGAACTCCAGAGGTGAAACCATTTGTTAAAAAGCCAGGATG 1200  
DB 1141 AATAGTATGAAAAAGAGAACTCCAGAGGTGAAACCATTTGTTAAAAAGCCAGGATG 1200  
QY 1201 GATTTATTTGCCAATGAACATCAGTGATTTCTCGA 1236  
DB 1201 GATTTATTTGCCAATGAACATCAGTGATTTCTCGA 1236

RESULT 4  
YSCNOFEAT  
LOCUS 1560 bp DNA linear PLN 27-APR-1993  
DEFINITION Saccharomyces cerevisiae transcription factor IIIA and RNA  
polymerase subunit RPB6 genes.  
ACCESSION M90638  
VERSION M90638.1 GI:172046

```

RESULT 5
YSCTFIIIA      1739 bp  DNA      linear  PLN 13-SEP-1996
LOCUS          Saccharomyces cerevisiae transcription factor IiIA (TFIIIA) gene,
DEFINITION     complete cds.
ACCESSION      M80611
VERSION        M80611.1  GI:172902
KEYWORDS       transcription factor IiIA; zinc-finger protein; zinc-finger
               transcription factor.
SOURCE         Saccharomyces cerevisiae (baker's yeast)
ORGANISM       Saccharomyces cerevisiae
               Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE      1 (bases 1 to 1739)
AUTHORS        A·chambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D.
               and Segall,J.
TITLE          The deduced sequence of the transcription factor TFIIIA from
               Saccharomyces cerevisiae reveals extensive divergence from Xenopus
               TFIIIA
JOURNAL        J. Biol. Chem. 267 (5), 3282-3288 (1992)
MEDLINE        92147684
PUBMED         1737784
FEATURES       Location/Qualifiers
               1..1739
               /organism="Saccharomyces cerevisiae"
               /mol_type="Genomic DNA"
               /strain="W303-1A"
               /db_xref="taxon:4932"
               /chromosome="XVI"
               234..1523
               /gene="TFIIIA"
               234..1523
               /gene="TFIIIA"
               /note="zinc finger containing protein"
               /codon_start=1
               /product="transcription factor IiIA"
               /protein_id="AAB08014.1"
               /db_xref="GI:172903"
               /translation="MGGEVLNNEGMPLAEIKQETPIRSSESSLSNSTSTSSSSN
               RPKTYGDYCDKAFTRPSILTEHOLSVHQGLRACDKCAKSVKSHLRHLVYH
               SDPKFQCGVCGKGVTRPQLKREHVHTKSFICPEGCNLRVFKHPOLRAHILSVHL
               HKLTCPNKNSFORPYLRNHIKSHDPEVENPYQCTFAGCKEERFWSLOLQSHKND
               HPKULCIPKCPGVENGLIHDDSLVTKNWKHCIFDMSFSRKHLLTHYSGI
               HTBEDIPLKIKYISDIQQIIVQDHGVGLSKNSHQDEDEKISNLRKRRRLTENNIV
               EFLQNEVDLEKRLSESGNGLNLLNTVGRKYRCFYNNKSTRTFFTKKEYKHDKHGK
               ELKLKILQEKSEENKTLVDQNHKEPFIQKTSAGDK"
ORIGIN
Query Match      17.6%; Score 217.4; DB 8; Length 1739;
Best Local Similarity 59.4%; Pred. No. 9.6e-26;
Matches 409; Conservative 0; Mismatches 261; Indels 9; Gaps 2;

Qy 62  GTCCCAAAAAGTATATTGCAATATGAAGGCTGTGATAAGCCATAATGACCATCAT 121
Db 367  GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAGACCTTCAA 426

Qy 122  TATTAGACACACATTTAAGACCCACACAGTAAATGATCGACCGTATAAATGTACAGTGGACG 181
Db 427  TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA--AGAGCATTTTCAGTGTGATA 483

Qy 182  ATTTGTGATAAAGCATTTTTCAGAAAAATCACATTTGGAAACACATATGTATGATCATTCGCG 241
Db 484  AGTGTGCAAAATCCCTTCGTTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCAATTCGT 543

Qy 242  AAAAAAACCATTCCATTGTTTCAGTGTCTGTGGTAAAGGGGTTAAATTCGCAACCACTTGA 301
Db 544  ATACGAACCATTCCAAATGTTCTTATTGTGGAAGAAGAGTGCAGCACTGCCAGCACTGA 603

Qy 302  AAAGACATGAAATCACCCATCAAAAGTCATTAAATGATFACATTTGAAATTTGTCAAGAAG 361
Db 604  AGCCACACGAAGTAACGATACCAAAATCTTTTCATTGTCCAGAGAAGAGTGCACACTCC 663

```

gene	CDS	/strain="S288C (AB972)" /db_xref="taxon:4932" /map="XVI" /note="baker's yeast" complement(12366..3655) /gene="TFC2" complement(12366..3655) /gene="TFC2" /note="p9677.9" /codon_start=1 /product="transcription factor TFIIA (PIR accession number S20050)" /protein_id="AAB64615.1" /db_xref="GI:786305"	GLQAYHACGRIBASIKVDFDKAIQRPSHTFAQYFKAISLQCDVGEVLESLELLEKVCQ EAAETESFQIGLVEIMRCSLDLXSQGLSKSIAKDTIERIKIISBELKCNQVW NYSQVLRFLWIKSVKVTDLVSVISFENSQFSBEIDSVDNITKTLDDSTDD NYSIAKTLILASKVSQDKFTDIAGTVRASVYNIIGISLTAFTITLKEPQYRDAAL PAFKSIQLOQNTSETLIGLATMDINFRVSOHCFIKATALEKATWNLWNLMLGL KKKDTFAQVNLKLSLAPODSSPWLGMALIEQGDIGSSKLFASFILSNGRSK AAQFPAQVNLHNNNGDDEDRDETVEKLTASTALEQFFKPSDSQALQCALLTL ERLHYENANELANLIGLKFKPKTQDERELFNPAIKQFAPARHILGLGNFELSIE NADLSQGISSESSDESKISNHCIGLSIFLNFDDQILNQFOELSSKSDSKHL VLLAKVLDVGSSTKBIALQELTEYATSGADLLVTLTAAMSILDDKRDLSIIL BELKALPLSKQIIDKHDAPLYLIEITKRLYRNDTGKQWORSAYFFNNLKVWERLD KNORRIASNGQNKVTAEEMSKLYCESKNLRSIQRMFLCPMNVTAVKALNCF"
		/gene="RPC3" complement(9710..11674) /gene="RPC3" /note="p9677.11" /codon_start=1 /product="74 kb subunit of RNA Polymerase III (Swiss Prot. accession number P32349)" /protein_id="AAB64619.1" /db_xref="GI:786307"	
gene	CDS	/gene="RPB6" join(3890..3909,3986..4433) /gene="RPB6" /note="p9677.8" /codon_start=1 /product="23 kb polypeptide of RNA Polymerases I, II, and III (Swiss Prot. accession number P20435)" /protein_id="AAB64616.1" /db_xref="GI:786304"	DJFLYKELVKAHLGERAASVIGMLVALGRLSVRELVKIDGMVDVSKITLVSLTLQR CVKYLQETAIISGKTYTYNEBGIHLLYSLIIDEITQMRVDEBEHKLQVAEIV QNVYSLGSLTVEDYLSVTSMSKTYISLFLQCEMGYLIQISKLHPTIPIEDLWQFL YEKYNIPRNSPLDLKKRSQAKNATDPAKINKPNLSQIILVDPKTSRLTVKP TVSLTINLDRFMKGRSKOLINLAKTRVSVTAQVYLAELRLAEALDRLGSLLRKSPDNKKRS LQDLBAKSFQDEALVEETPGLTFNAIDLAELPAELDLRGLSLLRKSPDNKKRS GSAASLPSKKLKTEDGVPALPAVAASKLSQESGDTQDEEDEDLADDPHSA SLINSHLKITASSNFFNETPGVYVYPSKMLPVLSVTEYVIASTLGSAMRLS RCIRNKLIVSEKINISTALMKEKDIRSLAIRYNSVEIQVPTADRSARAVFLP RCKETHSNFMQNLNWMANLLFKKELKQENSTLLKCANRDDVKRENNELLPSL NOLKMNRELRNFAVRLSRLSLWEVQMA"
		/translation="MSDIQEALSAENQGTGSEKLVTPEDVMTVISSLEQRTLNP DLFLYKELVKAHLGERAASVIGMLVALGRLSVRELVKIDGMVDVSKITLVSLTLQR CVKYLQETAIISGKTYTYNEBGIHLLYSLIIDEITQMRVDEBEHKLQVAEIV QNVYSLGSLTVEDYLSVTSMSKTYISLFLQCEMGYLIQISKLHPTIPIEDLWQFL YEKYNIPRNSPLDLKKRSQAKNATDPAKINKPNLSQIILVDPKTSRLTVKP TVSLTINLDRFMKGRSKOLINLAKTRVSVTAQVYLAELRLAEALDRLGSLLRKSPDNKKRS LQDLBAKSFQDEALVEETPGLTFNAIDLAELPAELDLRGLSLLRKSPDNKKRS GSAASLPSKKLKTEDGVPALPAVAASKLSQESGDTQDEEDEDLADDPHSA SLINSHLKITASSNFFNETPGVYVYPSKMLPVLSVTEYVIASTLGSAMRLS RCIRNKLIVSEKINISTALMKEKDIRSLAIRYNSVEIQVPTADRSARAVFLP RCKETHSNFMQNLNWMANLLFKKELKQENSTLLKCANRDDVKRENNELLPSL NOLKMNRELRNFAVRLSRLSLWEVQMA"	
gene	CDS	/gene="QCR2" 12014..13120 /gene="QCR2" 12014..13120 /note="p9677.6" /codon_start=1 /product="Ubiquinol-cytochrome C reductase core protein 2" /protein_id="AAB64620.1" /db_xref="GI:786302"	RIKETHSNFMQNLNWMANLLFKKELKQENSTLLKCANRDDVKRENNELLPSL NOLKMNRELRNFAVRLSRLSLWEVQMA"
		/translation="MLSAARLOFAGSRRRLTVSARDAPTSTKLSTLAVKVHGSRYATK DGAHLNLFNFQNTNRSALKLVRESSELLGTFKSTLDREVTILKANFLKDDLPVYV NALADVLYKTFAPHELTSLPAARYDVAECQPVKASQDLTAIFRKGSLNPLL YDCEVSVLQIDKFADKYTKENLEVENGEVYEAADLRFVDESLLSTLPAGKSLVSK SPKPFPLGNRVRFIGDSVAAIGIPVNAKSLAQVILANLYLSALSGLISSAKL DKFTDGLFTLFRDQDSAVSSNIKKIVADLKGKGLSPAINYTKLNAVQNESVSS PIELNPDAYDKFKLGFKNVAVGVDSNLPYLDEL" 14493..15410 /gene="p9677.5" 14493..15410 /gene="p9677.5" /note="Belongs to the MIP family of transmembrane channel proteins." /codon_start=1 /evidence=not experimental /protein_id="AAB64621.1" /db_xref="GI:786301"	
gene	CDS	/gene="p9677.12" complement(15542..16012) /gene="p9677.12" /note="Similar in C-terminus to Gentamicin 3'-acetyltransferase of P. aeruginosa (Swiss Prot. accession number P23181)"	complement(15542..16012) /gene="p9677.12" complement(15542..16012) /note="Similar in C-terminus to Gentamicin 3'-acetyltransferase of P. aeruginosa (Swiss Prot. accession number P23181)"

```
/codon_start=1
/evidence=not experimental
/protein_id="A864622.1"
/db_xref="GI:786308"
/translation="MNTSEDNITVRFVTENKQRLWKSYQDFYEVFFDDDDF
NFRFLDPIKMAVAVSSSEKIIMFNHMTWDFDKIYINDLYDENSRVK
GAGKLIQFYDEADKLGPVSVVWC'DESNHRQLLYVVKYKPKILYKRKY"
complement(16937..19570)
/genes="P9677.13"
/complement(16937..19570)
/genes="P9677.13"
/notes="Similar to S. cerevisiae hypothetical protein
HED799 (PIR accession number S45161) and S. pombe
17.6%; Score 217.4; DB 8; Length 37497;
Best Local Similarity 59.4%; Pred. No. 4.7e-26;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;
```

Qy 62 GTCCAAAAGATATTTGTCACATATGAAGGGTGTGATAAGCCTATATCGACCATCAT 121

Db 3522 GGCCAAAGACATATTTGTGACTATGATGGCTGTGATAAGGCATTTTCAAGACCTTCAA 3463

Qy 122 TATTAGAGCAACATTTAAGAACCCACAGTAATCATCGCGTATAAATGTACAGTGGACG 181

Db 3462 TTTTGACTGAACCAATTAAGCGGTACATCAGGGTTTA---AGAGCATTTTCAGTGTGATA 3406

Qy 182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAAACACATATTTGTATCACATTCG 241

Db 3405 AGTGTGCAAAATCCTTCGTTAAAGAGTCACTTAGAGAGACACTTGTATACGCAATTCG 3346

Qy 242 AAAAAAACCATTCCATTTGTCAGTGTGGTAAGGGTTAATCTCGACACACTTCA 301

Db 3345 ATAGAAACCATTCCTCAATGTTCTTATTGTGAAAAGAGTGAGCATCTCCAGCACTGA 3286

Qy 302 AAAGACATGAATCACCACATACCAATGATTTAAATGTACATTTGAAAAATTTGTCAGAG 361

Db 3285 AGCGACACGAAGTAAGCATACCAATCTTTTCATTTGTCAGAGGAAGATGCAACCTCC 3226

Qy 362 CATTTTAAACATCATCTTTTAAGACATCATATATATCTGTTCAATGAAGAAACATTTAA 421

Db 3225 GATTCTACAGCATCCACAAATTAAGGCGACATATTTTATCTGTCTATTTACATAAACA 3166

Qy 422 CGTGTAAACATGATATAAAGTTTTCACCTGACCTTCAAAATTAGCACACACATAAATTA 481

Db 3165 CTGTCCACACTGCAATAAAGCTTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 3106

Qy 482 AACATCA-----TGGTGGATCTCCCTGTTATCAATGTGATCATCTCGTGTGTTTAA 535

Db 3105 AACATCACGATCTCGAGTAGAAAATCCTTACCAGTGTACTTTTGTGCTGTGTTGCAAG 3046

Qy 536 ATTTCCAACTCGTGTGATTAACAATTCATATAAACAACATGCAATCCAAAACCTTAAT 595

Db 3045 AGTTTCGGATATGGTCACATTCGATTCGATATAAATAATGATCATCTTAATTAAT 2986

Qy 596 GTCTAAATGTGTAAGGTTGTGTTGGGAAAAAGGTTTATCTTCACATATGTTAAGTC 655

Db 2985 GTCTAATTTAGCAAAACCATGTGTGGGGAAAAATGGTTTACAAATGCACATGATTATTC 2926

Qy 656 ATGATGATTTACATGATCAAAATATGCACTTGTGATTAATTTGTGATGTTGGGAATTTG 715

Db 2925 ATGACGACTCACTAGTAACCAAAATTTGGAAGTGTATATATGTCCTGATATGCTCTTT 2866

Qy 716 CAAAGAAAAATGAATTAGTTGAACATTAT 744

Db 2865 CTAGAAAACATGATCTTCTCACGCATTAT 2837

RESULT 7

YSCRPO26A/c 2133 bp DNA linear, PLN 27-APR-1993

LOCUS S.cerevisiae RNA polymerase II sixth subunit (RP026) gene, complete cds.

DEFINITION YSCRPO26A

ACCESSION M33924

M33924.1 GI:172452

RNA polymerase II; c-myc proto-oncogene; transmembrane protein; tyrosine kinase

SOURCE Saccharomyces cerevisiae (baker's yeast)

ORGANISM Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 2133)

AUTHORS Archambault J., Schappert K.T. and Friesen J.D.

TITLE A suppressor of an RNA polymerase II mutation of Saccharomyces cerevisiae encodes a subunit common to RNA polymerases I, II, and III

JOURNAL Mol. Cell. Biol. 10 (12), 6123-6131 (1990)

MEDLINE 91061718

PUBMED 2247052

COMMENT Original source text: S.cerevisiae single-copy plasmid DNA, clone 2247052

2 Draft entry and computer-readable sequence for [Unpublished (1990)] kindly submitted by J.D.Friesen, 01-MAY-1990.

The Hospital for Sick Children

555 University Avenue

Toronto, Ontario, CANADA M5G 1X8

e-mail: jim@sickkids.toronto.edu.

Location/Qualifiers

1..2133

/organism="Saccharomyces cerevisiae"

/mol\_type="genomic DNA"

/db\_xref="taxon:4932"

join(1026..1045,1122..1569)

/note="RNA polymerase II sixth subunit (RP026)"

/codon\_start=1

/protein\_id="AAA34989.1"

/db\_xref="GI:172453"

/translation="MSDYEAFNDNGNENFEDVHFSDSEETBEKQFQDGETTDAN

GKTVTGGNGPDEQHEQIRRTKKEKAIKPKQDQATTPYMTKTERARILGTRALQIS

MNAVFVLDLGEETDPLRIKMLAEKKIPLVIRRYLPDGSFEDSVBELIVDL"

-1026..1045

/note="RNA polymerase II sixth subunit (RP026)"

/number=1

1046..1121

/note="RP026 intron A"

1122..1569

/note="RNA polymerase II sixth subunit (RP026)"

/number=2

Chromosome XVI.

ORIGIN

Query Match 16.1%; Score 199.2; DB 8; Length 2133;

Best Local Similarity 61.4%; Pred. No. 8e-23;

Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

Qy 171 TACAGTGGACGATTTGTGATAAGCATTTTTCAGAAAATCACATTTGGAACACATATTTGT 230

Db 556 TTCAGTGTGATGTTGTGCAAAATCTTCGTTAAAAGAGTCACTTAGAGACACTTGTGA 497

Qy 231 ATCACTTCGAAAAAACAACCATTCATTTGTTCAAGTGTGTTGTAAGGGGTTAATTTCTCG 290

Db 496 TACGCAATTCGATACGAAACCATTTCCAATGTTCTTATTGTGAAAAGAGTGACGACTCG 437

Qy 291 ACACACTTGAAGAGCATGAATCAACCCATACCAAGTCATTTAAATGTACATTTGAAA 350

Db 436 CCAGCACTGAAGCGACGACGAGTAACGATACCAATCTTTCATTTGTCCAGAGGAGG 377

Qy 351 TTGTCAAGAGCATTTTATAAATCAATCTTTTAAGACATCATATATATCTGTTTCATGA 410

Db 376 ATGCAACCTCGAATCTACAGCATCCCAATTAAGGCGACATATTTATCTGTTCAATTT 317

Qy 411 AABAACATTAAGGTGAACCAATGATTAAGTTTCTACTCGACCTTCAAAATTAGCACA 470

Db 316 ACATAAATAACCTGTCACACTGCAATAAAGCTTTTCAGAGGCCATATAGGCTAAGAA 257

Qy 471 ACATAAATAAACAATCA-----TGGTGGATCTCTCTGTTTATCAATGTGATCATCTCGG 524

```

Db      256 TCACATTTCTAAACATCAGATCCTGAGGTAGAAAAATCCTTACCAATGTACTTTTGCTGG 197
Qy      525 TTGTTTAAATTTCCAAATCTGGTCAGTATTACAAATTTTCATATAAACAACATGCGATCC 584
Db      196 TTGTTGCAAGAGTTTCCGATGATGTCACATTCGCAATTCGATATATAAATAATGATCATCC 137
Qy      585 AAAAATTTAAATGTCCTAAATGTTGTAAGGTGTTGGGAAAAAGTTTATCTTCACA 644
Db      136 TAAATTAATGTCCTATTGTTAGCAAAACCATGTTGTTGGGAAAAATGTTTACAAATGCA 77
Qy      645 TATGTTAAGTCATGATGATTCACATCATCAAAATATGGACTTGTGATATTGTTGATGT 704
Db      76 CATGATATTATTCATGACGACTCACTAGTAACCAAAAATTTGGAAGTGTATATATGTCCTGA 17
Qy      705 GGGGAAATTTGCAA 718
Db      16 TATGTCCTTTTCTA 3

CNS06ING 797 bp DNA linear STS 30-NOV-2001
LOCUS T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533
DEFINITION Saccharomyces bayanus, sequence tagged site.
ACCESSION AL400514.1 GI:12156635
VERSION Saccharomyces bayanus
KEYWORDS Saccharomycetes
SOURCE Saccharomycetes
ORGANISM Saccharomycetes
REFERENCE 1 (bases 1 to 797)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 797)
AUTHORS Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
Aigle,M. and Durrens,P.
Genomic exploration of the hemiascomycetous yeasts: 5.
Saccharomyces bayanus var. uvarum
FEBS Lett. 487 (1), 37-41 (2000)
20584715
PUBMED 11152880
REFERENCE 3 (bases 1 to 797)
AUTHORS Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Creneau, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1. .797
/organism="Saccharomyces bayanus"
/mol_type="genomic DNA"
/strain="CLIB 533"
/variety="uvarum"

FEATURES
source

```

```

/db_xref="taxon:4931"
/clone="AS0AA022B07"
/clone_lib="AS0AA"
/note="end : T7"
2. >796
/note="similar to Saccharomyces cerevisiae ORF YPR186c [
PF1 ; TFI1A (transcription initiation factor) ]"
/evidence=not_experimental

misc_feature
ORIGIN
Query Match 15.6%; Score 193; DB 11; Length 797;
Best Local Similarity 59.3%; Pred. No. 1e-21;
Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;
Qy 178 GACGATTTGATAAAGCATTTTTCAGAAATCACATTTTGGAAACACATATTGTATCATCAT 237
Db 83 GACATTTGTGCAAAATCCTTTGTTAARAAGAGTCATCTAGAGAGGCACCTGTTTTCGCAT 142
Qy 238 TCCGAAAAAACCATTCCATTGTTTCAGTGTGTGTAAGGGGTTAATTCTCGACAACAC 297
Db 143 TCTGATACGAAGCCATTCCTCAATGTTCTTATTTGTCAGAGGGGTGCAACTCGACAGCAA 202
Qy 298 TTGAAAGACATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAA 357
Db 203 TTGAAGCGCCACGAAGTTACATACCAATCTTTGTTTGTCCGGAAGATGGATGTGAT 262
Qy 358 GAAGCATTTTATAACATCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAACA 417
Db 263 CTCGGAATTTATAAGCACCTCAATTAAGGGTCTATATCTATCTGTCACCTAGACAA 322
Qy 418 TTAACCTGTAAACAATGTAAATTAAGTTTTCACTCGACCTTCAAAATTTAGCAACATATA 477
Db 323 TTAACCTGTCCCAATGTAGTAAATTTTTCAGAGACCTTACAGGCTAAGGAATCATATC 382
Qy 478 TTAACAATCATGTTGTT-----GATCTCTGTTTATCAATGTGATCATCTCGTTGTTT 531
Db 383 TCTAAACATCATGATCTGCTGAAGTGGTAAATCGGTATCAATGATCTTTCGCTGGCTGTT 442
Qy 532 AAAAATTTCCAAACTTGGTCAGTATTACAAATTTTCATATAAACAACATGATCCAAAAT 591
Db 443 ACAGAGTACCGTATATGTTGTCAGATGCAATGCAATACATATCAAAATGACCATCTAAAT 502
Qy 592 AAATGTCCTAAATGTGGTAAAGTTGTGTTGGGAAAAAGGTTTATCTTCATATGTTA 651
Db 503 AAATGCCCATTTGCAACAAGCTTTGTGAGGGGAAAAATGGTCTACAAATGATATGATC 562
Qy 652 AGTCATGATGATCTTCAACATGATCAAAATATGAGCTTGTGATTATTGATGTTGGGAAA 711
Db 563 ATTCAATGATGATCTTTGTTAAGAACTGGAATGTCATATTTGTTCTAATACATCT 622
Qy 712 TTTGCAAGAAAAATGAATTAGTTGAACATTAATATATCTTCCAT 756
Db 623 TTCTTGAGAAAGCACGACCTCTTACTCACTACCAAGATGACGAT 667

```

```

RESULT 9
AC138126/c 169063 bp DNA linear PRI 17-DEC-2002
LOCUS Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
DEFINITION AC138126
ACCESSION AC138126
VERSION AC138126.1 GI:27151357
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169063)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 169063)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
TITLE

```

JOURNAL Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
Location/Qualifiers  
1. .169063  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="RP11-274A19"

ORIGIN

Query Match 8.2%; Score 101; DB 9; Length 169063;  
Best Local Similarity 50.8%; Pred. No. 2e-07;  
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

QY 7 GAAAGTGAGCAACCAATCGATATCATCTTTAAATATCTTCTCTTCATCAGTCGCC 66

DB 80134 GAATGTGGCAAGCTTTTAAACMGTCCTCAATCCTTACTACACATAAGAAATTCATACT 80075

QY 67 AAAAAGTATATTTGCCATATGAAGGGTGTGATAAAGCCTAATAATCGACCAATCATATTA 126

DB 80074 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAGCTTTCTATCGATCTCTCAAAAGCT 80015

QY 127 GAGCAACATTTAGAACCCACACAGTATGATCGACCGTATAAATGTACAGTGGACGATTGT 186

DB 80014 ACTGACATTAAGAAATTTATCTGGAGAGAAACCTCACATGT-----GAAGAATGT 79861

QY 187 GATAAAGCATTTTTCAGAAAAATCACATTTGGAAAAACACATTTGTGTATCATCTCCGAAAAA 246

DB 79960 GGCAAAAGCCTTTAAACCATCTCTCACATCTTGTGTACACATAAGGTAAATTCATACGTGGAG 79901

QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGTTAATTCGACACACACATTTGAAAAAGA 306

DB 79900 AAACCTTACCAATGTGAGAAATGTGTAAAGCCTTTAACCGTCTCCACACCTTACTAGA 79841

QY 307 CATGAAATCACCATACAAAGTCATTTAAATGTACATTTGAAAAATGTCAAGACATTT 366

DB 79840 CATAAAGAGAATTCATCTGGAGAGAAACCCCTACCAATGTGAAAAATGTGGCAAGCTTTT 79781

QY 367 TATAACATCAATCTTTAA--GACATCATATATTATCTGTTTCATGAAAAACATTAAACG 423

DB 79780 AACAGTCTCAAACCTTTACTGGACATAGAAAAATTCATCTGGTGAGAACTCTACAA 79721

QY 424 TGTAAAACATGTATAAAGTTTTTCACTCGACCTTCAAAATTTAGCAACACATAAATTAATA 483

DB 79720 CCTAAAAGATGTAAACAGTGAATTTTGA AAAACACATTTCAAAGTTTCTTAAACATAAAAAAAT 79661

QY 484 CATCATGGTGGATCTCTCTGCTTATCAATGTGATCATCTCGTTGTGTTTTAAAAATTTCCAA 543

DB 79660 TATCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAAACAACCTTTAAAAAGTTGTAC 79601

QY 544 ACTTGGTCAGTATTACAATTTTCATATAAAAAACAACCTGCATCCAAAATCT 590

DB 79600 ACTTGAATGTGCAATAGATAAATTCATCTAAAAAAAACCTTCAAGT 79554

RESULT 10	AC073544/c	AC073544	177299 bp	DNA	linear	PRI 19-JUN-2002
LOCUS						
DEFINITION		Human sapiens chromosome 19 clone RP11-359H18, complete sequence.				
ACCESSION		AC073544				
VERSION		AC073544.4				
KEYWORDS		HTG.				
SOURCE		Human sapiens (human)				
ORGANISM		Human sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				

```

REFERENCE 1 (bases 1 to 177299)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177299)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 177299)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jun 19, 2002 this sequence version replaced gi:13699752.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
NOTE: BACTERIAL TRANSPOSONS excised at 14620 and 98099.
Location/Qualifiers
1. 177299
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-359H18"
FEATURES
source

```

ORIGIN					
	Query Match	8.2%	Score 101,	DB 9;	Length 177299;
	Best Local Similarity	50.0%;	Pred. No. 2e-07;		
	Matches 298;	Conservative	0;	Mismatches 280;	Indels 9; Gaps 2;
QY	7	GAAAGTCACGAACCAAAATCGATCATCATCTTTAATAATCTCTTCTCTCTTCCATCACGCTCCC	66		
Db	163740	GAATGTGGCAAGAGCTTTTAACCAAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT	163681		
QY	67	AAAAAGTATATTGCAACATATGAAGGGTGATAAAGCCTATTAATGCACCATCATTTATTA	126		
Db	163680	GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTCTATCGATCCTCAAAAACCT	163621		
QY	127	GAGCAACATTTAGAACCCACAGTAATGATCGACCCGTATAAATGTACAGTGGACGATGT	186		
Db	163620	ACTGAAATAGAANAATTCATCTGGAGAGAAACCCCTACACATGT-----GAAGAATGT	163567		
QY	187	GATAAAGCATTTTTGAGAAAATCACATTTGGAACAACATATTTGATCACATTCGAGAAAA	246		
Db	163566	GGCAAGGCCTTTAACCATTCCTCACATCTTGTACACATAAGSTAATTCATCTGGAGAG	163507		
QY	247	AAACCATTCATTTGTTCAAGTGTGTGGTAAAGGGTTAATTCCTGCACAACTTTGAAAAAGA	306		
Db	163506	AAACCCCTACCANAUTGAAGAATGTGGTAAAGCCCTTTAACAGTCTCACACCTTACTAGA	163447		
QY	307	CATGAATACCCATACAAAAGTCATTTTAAATGTACATTTGAAATTTCTCAAGAACATTTT	366		
Db	163446	CATAAGAGAAATTCATCTGGAGAGAAACCCCTACCAATGTGAAAAAATGTGGCAAGACTTTT	163387		
QY	367	TATAAAATCATCAATCTTTAA---GACATCATATATTATCTGTCATGAAAAAACATTAACG	423		
Db	163386	AACCAAGTCTCTCAAAACCTTTACTGGACATAAGAAAATTCATCTGGTGAAGAACTCTACAA	163327		
QY	424	TGTAACAATGTAATAAAGTTTTTCACCTCGACCTTCAAAATTAGCAACATTAATTAATA	483		
Db	163326	CCTAAAGATGTAACAGTGATTTTGTAAAAACACTTCAAAAGTTTTCTAAACATATAAAGAAAT	163267		
QY	484	CATCATGGTGAATCTCTGCTTATCATGTGATCATCTCGTGTGTTTTTAAAAATTTCCAA	543		
Db	163266	TATGCTGGTGAAGAAATCTTAGAATGTGAAGAATGTAAACAAACCTTTTAAAGTTGTAC	163207		
QY	544	ACTTGGTCAGTATTACAAATTTTCATATAAAAACAACTGCATCCAAAACCT	590		

QY	424	TGTAACCAATGTAATAAGTTTTCACCTCGACCTTCAAAATTTAGCACCAACATAAATTAAAA	483
Db	163326	CCTAAAGATGTAACAGTGATTTTGAACAACATTTTCAAAGTTTTCTAAACATAAAAAGAAAT	163267
QY	484	CATCAATGGTGAATCCTCGCTTATCAATGATGATCATCCTGGTGTGTTTTAAAAAATTTCCAA	543
Db	163266	TATGCTGGTGAGAATCTTAGAATGTGAAGAATGTACAAAACCTTTAAAAAGTTTGCAC	163207
QY	544	ACTTGGTCAGATTACAAATTTTCATATAAAACAACCTGCATCCAAAACCT	590

```

Db 163206 ACTTGATTGTCATAAGTAATTCATTACTATAAAAAAACCTACAAGT 163160

RESULT 11
AX714042
LOCUS AX714042 2132 bp DNA linear PAT 15-APR-2003
DEFINITION Sequence 726 from Patent EP1293569.
ACCESSION AX714042
VERSION AX714042.1 GI:29898970
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Iisogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,I., Otsuka,M., Nagahara,K. and
Masuho,Y.
Full-length cDNAs
Patent: EP 1293569-A 726 19-MAR-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
FEATURES
Location/Qualifiers
source 1..2132
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 8.0%; Score 99.4; DB 6; Length 2132;
Best Local Similarity 50.6%; Pred. No. 1e-06;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 7 GAAAGTACGAAACCAATCGATATCATCTTTTAATATCTTCTTCTTCATCAGTCCC 66
Db 1384 GAATGTGGCAAGCTTTTAACAGTCTCTCAATCTTACTACATAGAAGATTCATACT 1443
QY 67 AAAAAGTATATTTCACATATGAAGGCTGTGATAAGCCTATAATCGACCATCATATTA 126
Db 1444 GGAGAGAAATCTACAAATGTGAAGATGTGGCAAGCTTCTATCGATCTCAAACTT 1503
QY 127 GAGCAACATTTAGAACCCACAGTATGATCGACCGTAAATGATGACGTGACGATGT 186
Db 1504 ACTGAACTAAGAAATTTATCATCTGGAGAGAAACCTCTACACATGT-----GAAGAATGT 1557
QY 187 GATAAAGCATTTTTCAGAAATCACATTTTGGAAACACATATGTATCATCTCCGAAAAA 246
Db 1558 GGCAAGCCCTTTAACCACTCTCACACCTTGCTACACATAAGGTAATTCATCTGGAG 1617
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGTTAATTCGACACACATTTGAAAGA 306
Db 1618 AAACCCCTACCAATGTGAAGAAATGTGTAAGGCTTTTAACCACTCTCACACCTTACTAGA 1677
QY 307 CATGAATATCCCATACAAAGTCATTTAAATGTATCATTTTGAATAATGTCAAGACATTT 366
Db 1678 CATAAGAGAAATTCATCTGGAGAGAAACCTTACCAATGTGAATAATGTGGCAAGCTTTT 1737
QY 367 TATAACATCAATCTTTAA---GACATCATATATATATCTGTTTCATGAAAAACATTAACG 423
Db 1738 AACAGTCTTCAACCTTACTGACATAAGAAATTCATCTGGTGAAGAACTCTACAAA 1797
QY 424 TGTAACAATGTATAAAGTTTTCACCTCGACCTTCAAAATATAGCACACATAAATAA 483
Db 1798 CCTAAAAGATGTAAACAGTATTTTGAACACACTTCAAAGTTTCTAAACATAAAGAAAT 1857
QY 484 CATCATGTGTGATCTCTGCTTATCAATGTGATCATCTCGTGTGTTGTTTAAAAATTTCAA 543
Db 1858 TATGCTGGTGAGAAATCTTGAAGATGTGAAGATGTGAACAAACCTTTAAAGATTTGTCCAC 1917
QY 544 ACTTGTCAGTATTAACAATTTTATATAAACAACCTGCATCCAAAACCT 590
Db 1918 ACTTGATGTGCATAAGTAATTCATCTATAAAAAAACCTATAAGT 1964

```

## RESULT 12

AK056088

## LOCUS

AK056088

## DEFINITION

Homo sapiens cDNA FL31526 fis, clone NT2RI2000341, moderately similar to ZINC FINGER PROTEIN 43.

## ACCESSION

AK056088

## VERSION

AK056088.1 GI:16551397

## KEYWORDS

oligo capping; fis (full insert sequence).

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

1

## AUTHORS

Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furiya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshira,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Iisogai,T.

## TITLE

NEDO human cDNA sequencing project

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 2132)

## AUTHORS

Iisogai,T., Otsuki,T. and Sugiyama,T.

## TITLE

Direct Submission

## JOURNAL

Submitted (24-OCT-2001) Takao Iisogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

## COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- &amp; 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

## FEATURES

Location/Qualifiers

1..2132

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone\_lib="NT2RI2000341"

/cell\_line="NT2"

/cell\_type="teratocarcinoma"

/clone\_lib="NT2RI2"

/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"

148..1878

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAB71090.1"

/db\_xref="GI:16551398"

/translation="MVAFPPVCSHFQAFHSFQNIKDSFOKVPTRYGKCEHENLQL

SKSDEKVVQGGYNGLCPTTQSKIFQCDKYMKIHPFSLNGLGKVRTRKPKPK

YKEFGKFCIFNLTHQKLICTIRVNFYECDCGAFNGSSITFTKHIGHGKSYICE

EGKACNQFTNLTTHKIIITRDLYKRECSAFNLSHITHTIHTGPNYKREEC

DAFNQSTLTTHKIIITREKLYECCGKAFNOSHLTHKIIHTGKPYKCECGKAF

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

NKSLSLTHKIIHTGKPYKCECGKAFNOSHLTHKIIHTGKPYKCECGKAFN

QY 7 GAAAGTGACGAAACCAAAATGATATCATCTTTAAATATCTTCTTCTTCTTCTATCATCAGTCCC 66  
DB 1384 GAATGTGGCAAGCTTTTAAACCAAGTCTCTCAATCTTTACTACATAGAGAAATTCATCT 1443  
QY 67 AAAAAATATATTTGACATATGAAGGTTGATAAAGCCTTATAATCGACCATCATTAATTA 126  
DB 1444 GGAGAGAAATCTTACAAATGTGAAGATGTGGCAAGCTTTCTATCGATCTCTCAAACTT 1503  
QY 127 GAGCAACATTTAAGAACCCACAGATATGATCGACCGTATAAATGTACAGTGCAGCATTTG 186  
DB 1504 ACTGAACATAGAAGAAATTCATCTGGAGAGAAACCTCTACACATGT-----GAAGAATGT 1557  
QY 187 GATAAGCATTTTTCAGAAATACATTTTGGAAACACATATTTGATATCATCTTCCGAAAAA 246  
DB 1558 GGCAGAGCTTTTAAACCACTCTTACACCTTGTACACATAGATATTTCTATCTGGAGAG 1617  
QY 247 AAACCAATTCATTTTCAAGTGTGTGTAAGGGTTTAAATTTCTCGACAAACATTTGAAAGA 306  
DB 1618 AAACCCCTACCAATGTGAAGAAATGTGTAAAGCCTTTTAAACCAAGCTCTACACCTTACTAGA 1677  
QY 307 CATGAATACCCATACAAAGTCTTTAAATGTACATTTTGAATTTGTAAGTGTCAAGAGCATTT 366  
DB 1678 CATAAGAGAAATTCATCTGGAGAGAAACCCCTACCAATGTGAAGAAATGTGGCAAGCTTTT 1737  
QY 367 TATAAACATCAATCTTTTAA-----GACATCATATATTTCTTCTCATGAAAAAACATTAACG 423  
DB 1738 AACCAAGTCTCTCAACCTTACTGTGACATAGAAATTTCACTAGTGTGAGAACTCTACAAA 1797  
QY 424 TGTAAACAATGTATAAAGTTTTCATCTGACCTTCAAAATTTGACACAAACATTAATTAATA 483  
DB 1798 CCTAAAGATGTAAACAGTGTATTTGAAACACACTTCAAGTTTCTTAAACATTAAGAAAT 1857  
QY 484 CATCATGGTGCATCTCTGCTTATCAATGTGATCATCTCTGTTTGTATTAATAATTTCCAA 543  
DB 1858 TAGTCTGGTGAAGAAATCTTGAAGATGTGAAGATGTGAAGAAATTTTAAAGTTGTCTAC 1917  
QY 544 ACTTGTCAGTATTTACAAATTTTATATAAACAACACTGCATCCAAACT 590  
DB 1918 ACTTGATGTGCATAGATAATTCATCTACTAAAAAAACCTATAAGT 1964

RESULT 13  
AC099500/c 138627 bp DNA linear PRI 29-MAY-2002  
LOCUS Homo sapiens chromosome 19 clone RP11-209J6, complete sequence.  
DEFINITION AC099500  
ACCESSION AC099500.2 GI:21240690  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
www.jgi.doe.gov  
On May 29, 2002 this sequence version replaced gi:16930916.  
Draft Sequence Produced by DOE Joint Genome Institute  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;

Estimated Total Number of Errors is 0.1.  
NOTE: This insert is not the entire sequence of the clone (entire sequence is 153.8kb). It is clipped at the overlap with AC008981.  
The number of bases overlapped is 27192.  
Location/Qualifiers  
1. 138627  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="RP11-209J6"

FEATURES

source

ORIGIN

Query Match 8.0%; Score 99.4; DB 9; Length 138627;  
Best Local Similarity 50.6%; Pred. No. 3.8e-07;  
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;  
QY 7 GAAAGTGACGAAACCAAAATCGATATCATCTTTTAAATATCTTCTTCTTCTTCTATCATCAGTCCC 66  
DB 48699 GAATGTGGCAAGCTTTTAAACCAAGTCTCTCAATCTTTACTACATAGAGAAATTCATCT 48640  
QY 67 AAAAAATATATTTGACATATGAAGGTTGATAAAGCCTTATAATCGACCATCATTAATTA 126  
DB 48639 GGAGAGAAATCTTACAAATGTGAAGATGTGGCAAGCTTTCTATCGATCTCTCAAACTT 48580  
QY 127 GAGCAACATTTAAGAACCCACAGATATGATCGACCGTATAAATGTACAGTGCAGCATTTG 185  
DB 48579 ACTGAACATAGAAGAAATTCATCTGGAGAGAAACCTCTACACATGT-----GAAGAATGT 48526  
QY 187 GATAAGCATTTTTCAGAAATACATTTTGGAAACACATATTTGATATCATCTTCCGAAAAA 246  
DB 48525 GGCAGAGCTTTTAAACCACTCTTACACCTTGTACACATAGATATTTCTATCTGGAGAG 48466  
QY 247 AAACCAATTCATTTTCAAGTGTGTGTAAGGGTTTAAATTTCTCGACAAACATTTGAAAGA 306  
DB 48465 AAACCCCTACCAATGTGAAGAAATGTGTAAAGCCTTTTAAACCAAGTCTCTACACCTTACTAGA 48406  
QY 307 CATGAATACCCATACAAAGTCTTTAAATGTACATTTTGAATTTGTAAGTGTCAAGAGCATTT 366  
DB 48405 CATAAGAGAAATTCATCTGGAGAGAAACCCCTACCAATGTGAAGAAATTTTAAAGTTGTCTAC 48346  
QY 367 TATAAACATCAATCTTTTAA-----GACATCATATATTTCTTCTCATGAAAAAACATTAACG 423  
DB 48345 AACCAAGTCTCTCAACCTTACTGTGACATAGAAATTTCACTAGTGTGAGAACTCTACAAA 48286  
QY 424 TGTAAACAATGTATAAAGTTTTCATCTGACCTTCAAAATTTGACACAAACATTAATTAATA 483  
DB 48285 CCTAAAGATGTAAACAGTGTATTTGAAACACACTTCAAGTTTCTTAAACATTAAGAAAT 48226  
QY 484 CATCATGGTGCATCTCTGCTTATCAATGTGATCATCTCTGTTTGTATTAATAATTTCCAA 543  
DB 48225 TAGTCTGGTGAAGAAATCTTGAAGATGTGAAGATGTGAAGAAATTTTAAAGTTGTCTAC 48166  
QY 544 ACTTGTCAGTATTTACAAATTTTATATAAACAACACTGCATCCAAACT 590  
DB 48165 ACTTGATGTGCATAGATAATTCATCTACTAAAAAAACCTATAAGT 48119

RESULT 14

AC138469/c

LOCUS

DEFINITION

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AC138469 169500 bp DNA linear HTG 08-JAN-2003  
Homo sapiens chromosome 19 clone RP11-189C24, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 5 ordered pieces.  
AC138469  
AC138469.1 GI:27544967  
HTG; HTGS\_PHASE2; HTGS\_ACTIVEFIN.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 169500)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission

Unpublished  
2 (bases 1 to 169500)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Submitted (08-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 75797 75796: contig of 75796 bp in length  
\* 75797 75896: gap of unknown length  
\* 138072 138171: contig of 62175 bp in length  
\* 138172 157899: contig of 19728 bp in length  
\* 157900 157999: gap of unknown length  
\* 158000 166059: contig of 8059 bp in length  
\* 166059 166158: gap of unknown length  
\* 166159 169500: contig of 3342 bp in length.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 479013  
Center clone name: RP11-189C24  
-----  
Summary Statistics  
Consensus quality: 165586 bases at least Q40  
Consensus quality: 158891 bases at least Q30  
Consensus quality: 166156 bases at least Q20  
Estimated insert size: 16905000; agarose-fp estimation  
Estimated insert size: 166687; sum-of-contigs estimation  
Quality coverage: 0 in Q20 bases; agarose-fp estimation  
Quality coverage: 1 in Q20 bases; sum-of-contigs estimation.  
Location/Qualifiers  
1. 169500  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-189C24"  
-----  
ORIGIN  
Query Match 8.0%; Score 99.4; DB 2; Length 169500;  
Best Local Similarity 50.6%; Pred. No. 3.6e-07;  
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;  
QY 7 GAAAGTGACGAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66  
Db 49667 GAATGTGGCAAGCTTTTAAACCAAGTCTCAATCTTACTACACATAAGAGAAATCATACT 49608  
QY 67 ABAAGTATATTGACATATGAAGGTGTGATTAAGCCTATAATCGACCATCATATTA 126  
Db 49607 GGAGAGAAATCTACAAATGTGAGAAATGTGGCAAGCTTTCTATCATCTCTCAAACTT 49548  
QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCTGTAATGATGACGTGGACGATGT 186  
Db 49547 ACTGAACATAAGAAATTTATCTACTGGAGAGAAACCCCTACACATGT-----GAAGAAATGT 49494  
QY 187 GATAAGCATTTTTCAGAAATCACATTTTGGAAACACATATGTATCACATTCGGAATAA 246  
Db 49493 GGCAAGGCTTTTAAACCAATTCCTCACACCTTGTCTACATAAGTAATTCATCTACTGGAGNG 49434

QY 247 AAACCATTCATTTTTCAGTGTGTGTTAAAGGGTAAATTCCTCGACAAACATTTGAAAAGA 306  
Db 49433 AAACCCCTACCAATGGAAGAATGTGGTAAAGCCCTTAAACAGTCTCTCACACCTTACTAGA 49374  
QY 307 CATGAAATCACCACATACAAAGTCAATTTAAATGATACATTTGAAAATTTGCAAGAGCATTT 366  
Db 49373 CATAGAGAAATTCATACCTGAGAGAAACCCCTACCAATGTGAAAATTTGGCAAGCTTTT 49314  
QY 367 TATAACATCAATCTTTAA--GACATCATATATTTATGTTTCATGAAAACATTAACG 423  
Db 49313 AACCAAGTCTCTCAAACTTACTGACATAGAAAATTCATCTGGTGAGAACTTACAAA 49254  
QY 424 TGTAAACAATGTAATAAAGTTTTTCACTCGACCTTCAAAATAGCACAAACATTAATAA 483  
Db 49253 CCTAAAGATGTAACAGTGAATTTTGAACAACCTTCAAAAGTTTTTCTAAACATTAAGAAT 49194  
QY 484 CATCATGTTGGATCTCTCTCTTATCAATGTGATCATCTCGTGTGTTTAAAAATTTCCAA 543  
Db 49193 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAAATGTGAACAAAACCTTTAAAGTTGTCTAC 49134  
QY 544 ACTTGCTCAGTATTACAATTTTCATATAAACAACCTGCATCCAAAACCT 590  
Db 49133 ACTTGATTGTCATAAGATAATTCATCTATAAATAAACCCTATAAGT 49087  
RESULT 15  
AK122869  
LOCUS  
DEFINITION  
Homo sapiens CDNA FLJ16502 fis, clone FEBRA2006664, moderately similar to Zinc finger protein 43.  
ACCESSION  
AK122869.1 GI:34528067  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
AUTHORS  
Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2237)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
Location/Qualifiers  
1. 2237  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="FEBRA2006664"  
/tissue\_type="brain"  
/clone\_lib="FEBRA2"  
/dev\_stage="fetal"  
/note="cloning vector: pME18SFL3"  
FEATURES  
source

ORIGIN

```

Query Match      7.9%; Score 98.2; DB 9; Length 2237;
Best Local Similarity 52.0%; Pred. No. 1.6e-06;
Matches 273; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

QY 69 AAAGTATATTGCACATATGAGGGGTGATAAAGCCTATATCGACCATCATTTATTAGA 128
Db 1652 AGAGAAATCCTCAAAATGTGAGAAATGTGGCAAGCTTTCTATCGATCCTCAAACTTAC 1711
QY 129 GCAACATTTTAAGAACCCACAGTAATGATCGACCGGTATAAAATGTACAGTGGACGATTGTGA 188
Db 1712 TGAACATATAAGAAATTCATCTGGAGAGAAACCTTACACATGT-----GAAGAATGTGG 1765
QY 189 TAAAGCATTTTCAGAAATTCACATTTGGAACACATATTGTATCACATTCGAAAAAAA 248
Db 1766 CAAAGCCTTTAAACCATCTCTCACACCTTGTCTACACATAAGGTAATTCATCTGGAGAGAA 1825
QY 249 ACCATTCCCATTTGTTCAGTGTGTGTTAAAGGGTTAAATTCGACACACACTTGAAAAAGACA 308
Db 1826 ACCCTACCAATGTGAGAAATGTGTTAAAGCCTTTAACCAAGTCTCTCACACCTTACTAGACA 1885
QY 309 TGAATCACCCTACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCATTTTA 368
Db 1886 TAAAGAAATTCATCTGAGAGAAACCTTACCAATGTGAAAAATGTGGCAAGCTTTTAA 1945
QY 369 TAAACATCAATCTTTAA---GACATCATATATTATCTGTTTCAAGAAAAACATTTAAACGTG 425
Db 1946 CCAATCTCAAACTTACTGAGCATTAAGAAATTCATCTGTTGAGAACTCTACAACC 2005
QY 426 TAAACAAATGTAATAAGTTTTCATCTGACCTTCAAAATTTAGCACACATATAAATAAACA 485
Db 2006 TAAAGATGTACAGTGAATTTGAAAAACACTTCAAAGTTTCTAAACATATAAAGAAATTA 2065
QY 486 TCATGTTGATCTCTGCTTATCAATGTGATCATCTGCTGTTGTTTAAATTTCCAAAC 545
Db 2066 TGCTGTGAGAAATCTTAGAAATGTGAAGAAATGTAAACAAACCTTTAAAGTTGTCACAC 2125
QY 546 TTGGTCAGTATTACAATTTTCATATAAAACAACTGCATCCAAACT 590
Db 2126 TTGATTGTGCATTAAGATAATTCATATAAAAAAACCTTATAAGT 2170

```

Search completed: May 9, 2004, 08:56:45  
Job time : 5087.84 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:47:06 ; Search time 544.339 Seconds  
(without alignments)  
9646.138 Million cell updates/sec

Title: US-09-831-804-1\_COPY\_720\_1955

Perfect score: 1236  
Sequence: 1 atgagtgaagtgacgaac.....aaacatcagtgattcttcga 1236

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1236	100.0	2060	3	AA15398 DNA encod
2	1232.8	99.7	1239	6	AB232206 Candida a
3	99.4	8.0	2132	7	ADA53158 Human cod
4	96.8	7.8	1705	8	ACA98970 cDNA enco
5	96.6	7.8	2320	7	ACA56456 Human sig
6	95.2	7.7	4563	5	AAS91317 DNA encod
7	94.6	7.7	2597	7	AA151569 Human nuc
8	92.4	7.5	3639	5	AAS64586 DNA encod
9	91.8	7.4	2110	9	ADB63624 Human cDN
10	91.8	7.4	2760	9	ADC56695 Human mac
11	91.8	7.4	3078	9	ADB62468 Human cDN
12	90.4	7.3	2026	7	AAD55863 Human nuc
13	90.4	7.3	2114	7	ADA53124 Human cod
14	90.2	7.3	2230	7	ADA52931 Human cod
15	89.2	7.2	2084	9	ADC30762 Human nov
16	89.2	7.2	2622	5	AAS68872 DNA encod
17	89.2	7.2	2723	4	AAH16178 Human cDN
18	89.2	7.2	3839	6	ABK03826 Human cDN
19	89	7.2	2298	4	AAH16608 Human cDN
20	89	7.2	2905	5	AAS66143 DNA encod
21	89	7.2	3020	5	AAS92560 DNA encod
22	89	7.2	3502	4	AAI57845 Human pol
23	88.6	7.2	2597	7	AAD55855 Human nuc

24 88.6 7.2 4227 8 ACA98938  
25 88.4 7.2 1952 9 ADC58104  
26 88.4 7.2 2509 6 AAD31103  
27 87.6 7.1 1757 6 ABQ81135  
28 87.2 7.1 831 6 AAS55095  
29 87 7.0 976 5 AAS69188  
30 86.8 7.0 2662 6 ABQ93353  
31 86.8 7.0 6219 7 ACC46324  
32 86.8 7.0 6316 7 ACC46373  
33 86.2 7.0 1377 7 ABX34443  
34 86.2 7.0 2558 4 AAS26690  
35 86.2 7.0 2558 4 AAS26691  
36 86.2 7.0 2558 7 ABX74039  
37 86.2 7.0 2558 7 ABX74040  
38 86.2 7.0 2607 5 AAS87125  
39 86.2 7.0 2681 7 ABX34772  
40 86 7.0 1890 9 ADC30336  
41 86 7.0 3309 7 ACC46347  
42 86 7.0 3400 9 ADB62883  
43 85.6 6.9 1549 5 AAS67562  
44 85.6 6.9 2476 7 ACC46407  
45 84.8 6.9 2239 7 ADA53516

## ALIGNMENTS

RESULT 1  
AA15398  
ID AA15398 standard; DNA; 2060 BP.  
XX  
AC AA15398;  
XX  
DT 04-SEP-2000 (first entry)  
XX  
DE DNA encoding a transcription factor designated CATFIIA.  
XX  
KW Transcription factor; CATFIIA; DNA-binding protein;  
KW ribosomal RNA SS gene; fungal infection; ss.  
XX  
OS Candida albicans.  
XX  
FH Key Location/Qualifiers  
FT CDS 720..1958  
FT /\*tag= a  
FT /transl\_except= (pos: 1296..1298, aa: Ser)  
FT /transl\_except= (pos: 1734..1736, aa: Ser)  
XX  
PN WO200028037-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 09-NOV-1999; 99WO-FR002739.  
XX  
PR 10-NOV-1998; 98FR-00014147.  
XX  
PA (HMRI ) HOECHST MARION ROUSSEL.  
XX  
PI Bordon-Pallier F, Camier S, Sentenac A;  
XX  
WPI; 2000-376549/32.  
XX  
P-PSDB; AAY93316.  
XX  
PT New nucleic acid encoding Candida albicans transcription factor, useful  
XX e.g. in screening for antimycotic agents and for immunization.  
XX  
PS Claim 4; Page 32-33; 45pp; French.  
XX  
CC The present sequence encodes a Candida albicans transcription factor,  
XX designated CATFIIA. The polypeptide is a DNA-binding protein, which is  
XX involved in initiating transcription of the ribosomal RNA SS gene. The  
XX polynucleotide is used to screen for its specific inhibitors, potentially  
XX useful as antimycotic agents, to raise an antibody response that is

CC protective against fungal infection and to raise antibodies. Such  
CC antibodies, as well as the polypeptides and polynucleotides are used in  
CC compositions for diagnosing and treating fungal infections, e.g. by  
CC detecting polymorphisms and mutations  
XX  
SQ Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1236; DB 3; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 1e-233;  
Matches 1236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAGTGAAGGAGCAACCAACCAATCGATATCATCTTTAAATATCTCTCTCTTCATCA 60  
DB 720 ATGAGTGAAGGAGCAACCAACCAATCGATATCATCTTTAAATATCTCTCTCTTCATCA 779  
QY 61 COTCCCAAAAGATATATTTGACATATGAAGGTTGTATAAGCCTATAATTCGACCATCA 120  
DB 780 CGTCCCAAAAGATATATTTGACATATGAAGGTTGTATAAGCCTATAATTCGACCATCA 839  
QY 121 TTATTAGACACATTTAAGACCCACAGTAATGATCGACGTATAAATGTACAGTGGAC 180  
DB 840 TTATTAGACACATTTAAGACCCACAGTAATGATCGACGTATAAATGTACAGTGGAC 899  
QY 181 GATTGTGATAAGCATTTTTCAGAAAATCACAATTTGAAAACACATATTTGATCATCTCC 240  
DB 900 GATTGTGATAAGCATTTTTCAGAAAATCACAATTTGAAAACACATATTTGATCATCTCC 959  
QY 241 GAAAAAACCATTCCATTGTTTCAGTGTGTGTAAAGGGTTAATCTTCGACACACTTG 300  
DB 960 GAAAAAACCATTCCATTGTTTCAGTGTGTGTAAAGGGTTAATCTTCGACACACTTG 1019  
QY 301 AAAAGACATGAATCACCACATCAAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAA 360  
DB 1020 AAAAGACATGAATCACCACATCAAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAA 1079  
QY 361 GCATTTTAAACATCAATCTTTAAGACATATATATTCGTCATCGAAAAACATTA 420  
DB 1080 GCATTTTAAACATCAATCTTTAAGACATATATATTCGTCATCGAAAAACATTA 1139  
QY 421 ACCTGTAAACATGAATTAAGTTTTCCTCCACCTTCAAAATAGCACACATAAATTA 480  
DB 1140 ACCTGTAAACATGAATTAAGTTTTCCTCCACCTTCAAAATAGCACACATAAATTA 1199  
QY 481 AAACATCATGGTGATCTCTGCTATCAATGTGATCATCTCGTTGTTTAAAAATTC 540  
DB 1200 AAACATCATGGTGATCTCTGCTATCAATGTGATCATCTCGTTGTTTAAAAATTC 1259  
QY 541 CAAACTTGTGATGATTAACAATTTCAATATAAACAACCTGATCCAAACTTAAATGTCT 600  
DB 1260 CAAACTTGTGATGATTAACAATTTCAATATAAACAACCTGATCCAAACTTAAATGTCT 1319  
QY 601 AAATGTGTAAGGTTGTGTTGGGAAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660  
DB 1320 AAATGTGTAAGGTTGTGTTGGGAAAAAGTTTATCTTCACATATGTTAAGTCATGAT 1379  
QY 661 GATTCACCATCATCAAAATATGAGCTTGTGATTTATGATGTTGGGAAATTTGCAAG 720  
DB 1380 GATTCACCATCATCAAAATATGAGCTTGTGATTTATGATGTTGGGAAATTTGCAAG 1439  
QY 721 AAAAATGAATTAAGTTCGAACATTAATATCTTCCATGATGTTAATATCCCTGATGATTA 780  
DB 1440 AAAAATGAATTAAGTTCGAACATTAATATCTTCCATGATGTTAATATCCCTGATGATTA 1499  
QY 781 TTAAGGAACTGAAGTGAAGAAATAGAGACCTTATAGATCAAGGATCGAAATTAAT 840  
DB 1500 TTAAGGAACTGAAGTGAAGAAATAGAGACCTTATAGATCAAGGATCGAAATTAAT 1559  
QY 841 AATTTTCATGAATTAAGAAACAGAAATTAAGTGAAGAAAGATCAAGAGATCAAGAA 900  
DB 1560 AATTTTCATGAATTAAGAAACAGAAATTAAGTGAAGAAAGATCAAGAGATCAAGAA 1619  
QY 901 GATAGTCTAGATGAAGAAAGAGTGTATGATGATGATGATGATGATGATGATGATGAT 960

DB 1620 GATAGTCTAGATGAAGAAAGAGTGTATGATGATGATGATGATGATGATGATGATCA 1679  
QY 961 ATAAATCATTTACTGCTTCTTTGGAAGGTTCAAGAGTGTCTTAAACTTATTCGAAT 1020  
DB 1680 ATAAATCATTTACTGCTTCTTTGGAAGGTTCAAGAGTGTCTTAAACTTATTCGAAT 1739  
QY 1021 AGTGGGAAGAGATCAATTTGCTCTCTAAGATAATTTGTGATAGAAATGTTTCTAGAGAAAT 1080  
DB 1740 AGTGGGAAGAGATCAATTTGCTCTCTAAGATAATTTGTGATAGAAATGTTTCTAGAGAAAT 1799  
QY 1081 GATTTACGTCGACATTTGAAATGGCATGATATAATTTCAAGAAATTCAGTCATTCCTTA 1140  
DB 1800 GATTTACGTCGACATTTGAAATGGCATGATATAATTTCAAGAAATTCAGTCATTCCTTA 1859  
QY 1141 AATAGTATAGAAAAGAGAACTCCAGAGGTGACCACTTGTGTTAAAAAGCCAGGATG 1200  
DB 1860 AATAGTATAGAAAAGAGAACTCCAGAGGTGACCACTTGTGTTAAAAAGCCAGGATG 1919  
QY 1201 GATTTATTCGCAATGAAACATCAGTGATTTCTCGA 1236  
DB 1920 GATTTATTCGCAATGAAACATCAGTGATTTCTCGA 1955  
RESULT 2  
ABZ32206  
ID ABZ32206 standard; DNA; 1239 BP.  
XX ABZ32206;  
AC ABZ32206;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
XX Candida albicans essential gene SEQ ID NO 6493.  
DE Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.  
XX  
OS Candida albicans.  
XX  
XX WO200253728-A2.  
XX  
XX 11-JUL-2002.  
XX  
XX 26-DEC-2001; 2001WO-US049486.  
XX  
XX 29-DEC-2000; 2000US-0259128P.  
PR 20-FEB-2001; 2001US-00792024.  
PR 22-AUG-2001; 2001US-0314050P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
XX WFI; 2002-566694/60.  
DR P-PSDB; ABP73656.  
XX  
XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele of  
PT a gene and placing other allele of the gene under conditional expression.  
XX  
XX Claim 37; SEQ ID NO 6493; 167pp + Sequence Listing; English.  
PS  
XX The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene

CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthesis, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans gene used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office  
XX  
SQ Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;  
Query Match 99.7%; Score 1232.8; DB 6; Length 1239;  
Best Local Similarity 99.8%; Pred. No. 3.9e-233;  
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGAGTGAAGTGCAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60  
DB 1 ATGAGTGAAGTGCAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60  
QY 61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAGCCCTATATCGACCATCA 120  
DB 61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAGCCCTATATCGACCATCA 120  
QY 121 TTATTAGAGCAACATTTAAGAACCCACAGTAAGTGCAGCCGTATAAATGTACAGTGGAC 180  
DB 121 TTATTAGAGCAACATTTAAGAACCCACAGTAAGTGCAGCCGTATAAATGTACAGTGGAG 180  
QY 181 GATTGTGATGAAGCATTTTTCAGAAATCACATTTGGAAACACATATTTGTATCACATTC 240  
DB 181 GATTGTGATGAAGCATTTTTCAGAAATCACATTTGGAAACACATATTTGTATCACATTC 240  
QY 241 GAAAAAACAACATTTCCATTTGTCAGTGTGTGTAAGGGGTAAATCTCGACACACTTG 300  
DB 241 GAAAAAACAACATTTCCATTTGTCAGTGTGTGTAAGGGGTAAATCTCGACACACTTG 300  
QY 301 AAAAGACATGAATCAACCCATACAAAGTCAATTTAAATGTACATTTGAAATTTGCAAGAA 360  
DB 301 AAAAGACATGAATCAACCCATACAAAGTCAATTTAAATGTACATTTGAAATTTGCAAGAA 360  
QY 361 GCATTTTATAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTA 420  
DB 361 GCATTTTATAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAACATTA 420  
QY 421 ACGTGTAAACATGTAATGAAGTTTTCACCTGACCTTCAAAATTAGCACACATTAATTA 480  
DB 421 ACGTGTAAACATGTAATGAAGTTTTCACCTGACCTTCAAAATTAGCACACATTAATTA 480  
QY 481 AAACATCATGTGGATCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAATTC 540  
DB 481 AAACATCATGTGGATCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAATTC 540  
QY 541 CAAACTTGGTCAGTATTACAAATTTTCATATAAACAACCTGCATCCAAACTTAATGTCCT 600  
DB 541 CAAACTTGGTCAGTATTACAAATTTTCATATAAACAACCTGCATCCAAACTTAATGTCCT 600  
QY 601 AAATGTGTAAGGTTGTGTTGGGAAAAAGGTTTATCTTCACATATGTTAAAGTCATGAT 660  
DB 601 AAATGTGTAAGGTTGTGTTGGGAAAAAGGTTTATCTTCACATATGTTAAAGTCATGAT 660  
QY 661 GATTCTACCATGATCAAAATATGACCTTGTGATTTATGTTGTTGGGAAATTTGCAAG 720  
DB 661 GATTCTACCATGATCAAAATATGACCTTGTGATTTATGTTGTTGGGAAATTTGCAAG 720  
QY 721 AAAAATGAATGTTGACATTAATATCTTCATGATGTAATATCCCTGATGATTTA 780  
DB 721 AAAAATGAATGTTGTTGAACATTAATATCTTCATGATGTAATATCCCTGATGATTTA 780

QY 781 TTAAGGAAACTGAAGTGAAAAAATTTAGAGAACCTATTAGATCAAGGATCGAAATTAAT 840  
DB 781 TTAAGGAAACTGAAGTGAAAAAATTTAGAGAACCTATTAGATCAAGGATCGAAATTAAT 840  
QY 841 AATTGTCATGAATTTAGAAAAACAGAGAAATTTAAAGTGAAGAGATGAAGATGAAGAA 900  
DB 841 AATTGTCATGAATTTAGAAAAACAGAGAAATTTAAAGTGAAGAGATGAAGATGAAGAA 900  
QY 901 GATAGTCTAGATGAAAAAAGAGAGTGTGTAGATCAGACTCAATGTGAGCTCAAGATCA 960  
DB 901 GATAGTCTAGATGAAAAAAGAGAGTGTGTAGATCAGACTCAATGTGAGCTCAAGATCA 960  
QY 961 ATAAAAATCATTTACTGCTTCTTCTGGAAGGTTTCAAGAGTGTCTTAACTTATCTGAT 1020  
DB 961 ATAAAAATCATTTACTGCTTCTTCTGGAAGGTTTCAAGAGTGTCTTAACTTATCTGAT 1020  
QY 1021 AGTGGGAAGAGATCAATTTGCTTAAGAAATTAATTTGTGATGAATGTTTTCTAGAGATAT 1080  
DB 1021 AGTGGGAAGAGATCAATTTGCTTAAGAAATTAATTTGTGATGAATGTTTTCTAGAGATAT 1080  
QY 1081 GATTACCTGACATTTGAAATGGCATGATGATAATTTACAAAGATTTGAGTCAATCTTA 1140  
DB 1081 GATTACCTGACATTTGAAATGGCATGATGATAATTTACAAAGATTTGAGTCAATCTTA 1140  
QY 1141 AATAGTATAGAAAAAGAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGGATG 1200  
DB 1141 AATAGTATAGAAAAAGAGAACTCCAGAGGTGAACCATTTGTTAAAAAGCCAGGATG 1200  
QY 1201 GATTTATTGCCAAATGAAACATCATGATTTCTCGA 1236  
DB 1201 GATTTATTGCCAAATGAAACATCATGATTTCTCGA 1236  
RESULT 3  
ADA53158  
ID ADA53158 standard; cDNA; 2132 BP.  
XX ADA53158;  
XX AC ADA53158;  
XX DT 20-NOV-2003 (first entry)  
XX DE Human coding sequence, SEQ ID 726.  
XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
XX KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
XX KW inflammatory disease; osteoporosis; neurological disease; gene; ss.  
XX OS Homo sapiens.  
XX PN EP1293569-A2.  
XX PD 19-MAR-2003.  
XX PF 21-MAR-2002; 2002EP-00006586.  
XX PR 14-SEP-2001; 2001JP-00328381.  
XX PR 24-JAN-2002; 2002US-0350435P.  
XX PA (HELI-) HELIX RES INST.  
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuho Y;  
XX WPI: 2003-395539/38.  
XX P-PSDB; ADA54797.  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX Claim 1; SEQ ID NO 726; 205pp; English.  
PS

```

XX CC The present invention relates to novel human secretory or membrane
XX CC proteins (ADA54072-ADA54071) and their coding sequences (ADA52433-
XX CC ADA54071). The coding sequences are useful in the gene therapy of
XX CC diseases caused by abnormalities of the proteins, e.g. cancer,
XX CC inflammatory diseases, osteoporosis or neurological disease.
XX CC
XX CC Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;
XX CC
XX CC Query Match 8.0%; Score 99.4; DB 7; Length 2132;
XX CC Best Local Similarity 50.6%; Pred. No. 3.4e-10;
XX CC Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
XX CC
XX CC QY 7 GAAAGTGCAGAAACCAATCGATATCACTTTAATATCTCTTCTTCATCAGGCC 66
XX CC Db 1384 GAATGTGCAAGCTTTTAAACAGTCTCAATCTTACTACATAGAGAAATCACT 1443
XX CC
XX CC QY 67 AAAAAGTATATTGCACATATGAAGGGTGTGATAAAGCCTAATACGACCATCATTA 126
XX CC Db 1444 GGAGAGAAATCTACAAATGTGAAGATGTGGCAAGCTTTCTATCGATCTCAAAACTT 1503
XX CC
XX CC QY 127 GAGCAGACATTAAGAACCCACAGTAATGATCGACCGTATTAATGATGAGCGATTGT 186
XX CC Db 1504 ACTGAACATAGAAAATTCATCTCGAGAGAAACCCCTACACATGT-----GAAGAATGT 1557
XX CC
XX CC QY 187 GATAAAGCATTTTTCAGAAATTCACATTTGGAAACACATATTTGATCAGATTCGAAAAA 246
XX CC Db 1558 GCGAAGCCTTTAACACCTCTCCACCTTGCTACACATAGGTAATTCATCTGGAGAG 1617
XX CC
XX CC QY 247 AAACCATTCATTTGTCAGTGTGTAAGGGGTTAATCTCGACACACATCGAAAGA 306
XX CC Db 1618 AAACCTTACCAATGTGAAGATGTGTAAGGCCCTTTAACCGATCTCCACACCTTACTAGA 1677
XX CC
XX CC QY 307 CATGAAATCACCACACAAAGTCATTAAATGTACATTTGAAATTTGCAAGAGCATTT 366
XX CC Db 1678 CATAGAGATTCATCTCGAGAGAAACCTTACCATGTGAAATGTGGCAAGCTTTT 1737
XX CC
XX CC QY 367 TATAAATCAATCTTTAA---GACATCATATATATCTGTTCTATGAAAAAACHTTAAG 423
XX CC Db 1738 AACCAAGTCTCAAACTTACTGGACATAGAAATTCATCTGTTGAGAACTCTCAAA 1797
XX CC
XX CC QY 424 TGTAAACATGTAATAAGTTTTCCTCGACTTCAAAATAGACACACATATAAATAA 483
XX CC Db 1798 CTTAAAGATGTAACAGTGTATTTGAAACACTTCAAGATTTCTTAACATAGAAAT 1857
XX CC
XX CC QY 484 CATCATGTGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTGTTTAAATTTCCAA 543
XX CC Db 1858 TATGCTGTGAGAAATCTTAGAAATGTGAAGATGTGAAGAACTTAAAGTTGTAC 1917
XX CC
XX CC QY 544 ACTTGTCAGTATTACATTTTCATATAAACAACCTGCACTCCAAACT 590
XX CC Db 1918 ACTTGATTTGTGATAAGATAATTCATCTACTTAAATAAATAAACCCTATAAGT 1964
XX CC
XX CC RESULT 4
XX CC ID ACA98970
XX CC AC ACA98970;
XX CC XX ACA98970;
XX CC DT 25-JUL-2003 (first entry)
XX CC XX
XX CC DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX CC KW Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
XX CC KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
XX CC KW anti-allergic; anti-inflammatory; thyromimetic; gene therapy;
XX CC KW cell proliferative disorder; cancer; atherosclerosis;
XX CC KW neurological disorder; epilepsy; Huntington's disease; stroke;
XX CC KW immune disorder; inflammatory disorder; AIDS; allergy;
XX CC KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
XX CC KW protein-protein interaction; drug-target interaction;
XX CC KW gene expression profile; gene; ss.

```

```

XX OS Homo sapiens.
XX PN WO2003023003-A2.
XX PD 20-MAR-2003.
XX PP 05-SEP-2002; 2002WO-US028540.
XX PR 07-SEP-2001; 2001US-0317792P.
XX PR 07-SEP-2001; 2001US-0317912P.
XX PR 14-SEP-2001; 2001US-0322270P.
XX PR 21-SEP-2001; 2001US-0324040P.
XX PR 28-SEP-2001; 2001US-0326732P.
XX PR 19-OCT-2001; 2001US-0346716P.
XX PR 25-JAN-2002; 2002US-0351749P.
XX PR 22-FEB-2002; 2002US-0359498P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX XX
XX PI Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ;
XX PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BM;
XX PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
XX PI Hafalia AJA, Sanjanwala B, Marguis JP, Gorvad AP, Lee SY, Ison CH;
XX PI Baughn WK, Chawla NK, Nguyen DE, Swarnakar A, Zebarjadian Y, Shah P;
XX PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE;
XX PI Burford N, Ramkumar J;
XX XX
XX WPI: 2003-313243/30.
XX P-PSDB; ABU96722.
XX
XX PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
XX PT treating and preventing diseases or conditions associated with the
XX PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
XX PT infections.
XX PS Claim 5; Page 340-341; 345pp; English.
XX
XX CC The invention describes a novel human isolated nucleic acid-associated
XX CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
XX CC diagnosing, treating and preventing diseases or conditions associated
XX CC with the decreased expression or overexpression of NAAP, such as cell
XX CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
XX CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
XX CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
XX CC disorders, or infections. These are also useful in assessing the effects
XX CC of exogenous compounds on the expression of nucleic acid and amino acid
XX CC sequences of NAAP. The NAAP or its fragments are useful in screening
XX CC compounds for effectiveness as agonist or antagonist of the polypeptides,
XX CC or in altering the expression of the target polynucleotide and compounds
XX CC that specifically bind to or modulate the activity of the polypeptide.
XX CC The microarray is useful in monitoring or measuring protein-protein
XX CC interactions, drug-target interactions, and gene expression profiles.
XX CC This sequence encodes a novel human nucleic acid-associated protein
XX CC (NAAP).
XX SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

```

```

XX SQ
XX Query Match 7.8%; Score 96.8; DB 8; Length 1705;
XX Best Local Similarity 52.9%; Pred. No. 1.1e-09;
XX Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;
XX
XX QY 7 GAAAGTGCAGAAACCAAAATCGATATCATCTTTAATATCTTCTTCTTCATCAGCTCCC 66
XX Db 695 GAATGTGCGAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAGAAATCCAAACT 754
XX
XX QY 67 AAAAAGTATATTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTTATA 126
XX Db 755 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCCTTTAACTGGTCCCAACCCCT 814
XX
XX QY 127 GAGCAACATTTAAGAACCCACAGTATGATCGACCGTATATAATGTACATGACGATGTGT 186
XX Db 815 AATAAACAATAAGAGATTTCATCTACTGAGAAAAACCTTACAAATGTAAA-----GAATGT 868

```

187 GATAAAGCATTTTTCAGAAATACATTTTGGAAACATATTTGTATCATCTCCGAAAA 246  
869 GGCAGAGCCTTTAAACAGAGCTTACACCTTTATTAGACATAAGAGAAATTCATCTGAAGAG 928  
247 AAACCATTCATTTTCAGTGTGTGTAAAGGGTTAATTTCTGACACACCTTGAAGA 306  
929 AAACCCATCAAAATGTGAAGAATGTGCAAGGCTTTTAAACAGTCTACGACCTTACTACA 988  
307 CATGAATACCCATACAAAGTCATTTAAATGTACATTTTGAATAATTTGCAAGAACATTT 366  
989 CATAATATATTCATCTGGGAAATTCCTACAAATGTGAGAAATGTGTAGAGCTTT 1048  
367 TATAACATCAATCTTTAAGACATCATATTTATCTGTT---CATGAAAAACATTAAGC 423  
1049 AACCAAGCCTCAAGCTTTACTGAACATAAGTTAATTCATACCGGAGAGAAATGTCTACAA 1108  
424 TGTAAACAAATGTAATAAGTTTTTCACTCGACCTTCAAAATTTAGCAACACATAAATTTAAA 483  
1109 TGTGAAGATGTGGCAAGTTTTTAAACCGATCTTCAACCTTACAGACCAATAAAAAATT 1168  
484 CATCATGGTG 493  
1169 CATACAGGAG 1178

RESULT 5

ACA56456  
ID ACA56456 standard; cDNA; 2320 BP.

ACA56456;

06-JUN-2003 (first entry)

Human signalling pathway polynucleotide probe SEQ ID NO 1054.

Human; probe; ss; array element; Parkinson's disease;  
signalling pathway population; cancer; adenocarcinoma; leukaemia;  
immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.

Homo sapiens.

US6500938-B1.

31-DEC-2002.

30-JAN-1998; 98US-00016434.

30-JAN-1998; 98US-00016434.

(INCY-) INCYTE GENOMICS INC.

Au-Young J, Seilhamer JJ;

WPI; 2003-352189/33.

Combination of polynucleotide probes, useful as array elements in a  
microarray for monitoring the expression of a number of target  
polynucleotides.

Claim 1; SEQ ID NO 1054; 65pp; English.

The invention relates to a combination which, comprises a number of  
polynucleotide probes comprising a sequence selected from one of the 1490  
sequences mentioned in the specification. The combination is useful as an  
array element in a microarray for monitoring the expression of a number  
of target polynucleotides. The microarray is particularly useful in the  
diagnosis and treatment of cancer and immunopathology and neuropathology.  
The microarray is useful in diagnostics and treatment regimens, drug  
discovery and development, toxicological and carcinogenicity studies,  
forensics and pharmacogenomics. The microarray is also useful for  
monitoring progression of diseases and for developing sophisticated  
profiles for the effects of currently available therapeutic drugs. The

combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
and genomic fragments and in research and diagnostic applications. The  
array can detect changes in expression in a large number of genes coding  
for different signaling pathway populations which can be used to diagnose  
various diseases including cancer e.g. adenocarcinoma and leukaemia,  
immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease  
and Parkinson's disease. The present sequence represents a polynucleotide  
probe of the invention. Note: The sequence data for this patent did not  
form part of the printed specification but was obtained in electronic  
format directly from USPTO at  
seqdata.uspto.gov/sequence.html?docID=06500938B1

Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.6; DB 7; Length 2320;  
Best Local Similarity 51.3%; Pred. No. 1.2e-09;  
Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGATATATTGACATATGAGGGTGTATTAAGCCTATATCGACCATCATTTATAGA 128  
Db 964 AGAGAAACCCCTACAAATGTAAAGAAATGTGTAAGAGCTTTTAAACCGATCTTCAACCTTAC 1023  
QY 129 GCAACATTTAGAACCCACAGTATGATCGACCGCTATATAATGTACAGTGGACGATTTGA 188  
Db 1024 TACCATAGAAAAATTCATCTGAGAGAAACCTTACAAATGT-----GAAGAATGTGG 1077  
QY 189 TAAAGCATTTTTCAGAAAAATCACAATTTGGAAACACATATTGTATCATCATCTCCGAAAAAA 248  
Db 1078 CAAAGCCTTTAAGCAGTCTCTCAAAACCTTACTACACATAAGATAATTCATCTGAGAGAA 1137  
QY 249 ACCATTCCATTTTTCAGTGTGTGTAAGGGGTTAATTTCTCGACACACATTTGAAAGACA 308  
Db 1138 ACCCTACAAATGTAAAGAAATGTGGAAGCCCTTTAAACAGTCTGACACCTTACACACA 1197  
QY 309 TGAATACCCCATACAAAGTCATTTAAATGTACATTTTGAATAATGTCAAGAGCATTTTA 368  
Db 1198 TGAGTAAATTCATCTGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCCTTTAA 1257  
QY 369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTCATCGAAAAACATTAACGTG 425  
Db 1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATCTGAGAGAAACCTTACAAATG 1317  
QY 426 TAAACAATGTAATAAGTTTTTCACTCGACCTTCAAAATTTAGCAACACATAATAATAACA 485  
Db 1318 TAAAGATGTGTAAGCTTTTAAACACTCTTCAACCTTACTAAACATAAGATAATTA 1377  
QY 486 TCATGGTGATCTCTGCTTATCAATGTGATCATCTCGTGTGTTTAAAAATTTCCAAAC 545  
Db 1378 TACTGGAGAGAGCCCTTACAAATCTAAAGAAATGTGAAAAAGCTTTTAAACCAATCCTCAA 1437  
QY 546 TTGGTCAGTATTACAAATTTTCAATAAACAACCTGATCCAAACTTAAATGTCTTAATG 605  
Db 1438 ACTTACTGAACATAAGAAAAATTCATCTGGAGAGAAACCTTA---TGAATGTGAAAAATG 1494  
QY 606 TGGTAAAGGTGTGTTGGGAAAAAGGTTTATCTTTCATATATGTTAGTCAATGATGA 662  
Db 1495 TGGCAAGGTTTAAACCAAGTCTCTCAAAATCTTACTAGACATAAGAAAAAGTCAACAGA 1551

RESULT 6

AAS91317  
ID AAS91317 standard; cDNA; 4563 BP.

AC AAS91317;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #27121.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.  
Homo sapiens.

XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 XX 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX P-PSDB; ABG27130.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity.  
 XX Claim 1; SEQ ID NO 27121; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 XX sequences, (I) is useful as hybridisation probes, polymerase chain  
 XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 XX and in recombinant production of (II). The polynucleotides are also used  
 XX in diagnostics as expressed sequence tags for identifying expressed  
 XX genes. (I) is useful in gene therapy techniques to restore normal  
 XX activity of (II) or to treat disease states involving (II). (II) is  
 XX useful for generating antibodies against it, detecting or quantitating a  
 XX polypeptide in tissue, as molecular weight markers and as a food  
 XX supplement. (II) and its binding partners are useful in medical imaging  
 XX of sites expressing (II). (I) and (II) are useful for treating disorders  
 XX involving aberrant protein expression or biological activity. The  
 XX polypeptide and polynucleotide sequences have applications in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits to assess biodiversity  
 XX and to produce other types of data and products dependent on DNA and  
 XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 XX coding sequences of the invention. Note: The sequence data for this  
 XX patent did not appear in the printed specification, but was obtained in  
 XX electronic format directly from WIPO at  
 XX fip.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;  
 Query Match 7.7%; Score 95.2; DB 5; Length 4563;  
 Best Local Similarity 52.7%; Pred. No. 2.5e-09;  
 Matches 256; Conservative 0; Mismatches 223; Indels 9; Gaps 2;  
 QY 7 GAAAGTCAGCAACCAACGATATCATCTTTAATATCTTCTTCTTCATCGTCCC 66  
 DB 1315 GAATGTGGGAAAGCCTTTAACCGTCTCTACACCTTACTCAACATAAACCAATTCTACT 1374  
 QY 67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAGCCTATTAATCGACCATCTATTATTA 126  
 DB 1375 GGAGGGAACACCTACAAATGTGAAGAAATGTGGCAAGCTTTTAAACATCTTTCAGCCCTT 1434  
 QY 127 GAGCACAATTTAGACACCCACAGTATATGATCGACGCTATATGATGACAGTGACGATGT 186  
 DB 1435 ACTAACATTAAGATATTTATCTATCGGGGAAACCATCAATATG-----GAGAAATGT 1488  
 QY 187 GATAAAGCATTTTTCAGAAAATCACAATTTGGAACACATATTGTATCATCTCCGAAAAA 246  
 DB 1489 GGCAAAAGCTTTTAGGCAATCTCTCACACCTTACTAGACATAAAGCAATTCATCTGGAGAG 1548  
 QY 247 AAACCATTCATTTTCAGTGTGTGTAAGGGTAAATTCGACACACATCTGAAAGA 306  
 DB 1549 AAACCATTCATTTTCAGTGTGTGTAAGGGTAAATTCGACACATCTGAAAGA 1608  
 QY 307 CATGAAATCACCATCAACAGTCAATTTAAATGATACATTTGAAATTTGCAAGAGCATTT 366

DB 1609 CATAAGATAATTCATCTGGAAGAAACCCCTACAATGTGAAGATGTGGGAAGCTTTT 1668  
 QY 367 TATAACATCAATCTTTTAAAGA---CATCATATATTATCTGTTCATGAAAAACATTAACG 423  
 DB 1669 AGCCAGTCTCTCAACCCCTTAGAAAACCATCAGATAAATTCATCTGGAGAGAAACCCCTACAAA 1728  
 QY 424 TGTAAACAATGTAATAAAGTTTTTCTACTGACCTTCAAAATTAGCACACATATAATTTAAAA 483  
 DB 1729 TGTGAAGAATGTGGTAAAGCTTTTAAAGTGGTTCATCAAAACTTACTGTACATAAGGTAATT 1788  
 QY 484 CATCATGGTG 493  
 DB 1789 CATACTGGAG 1798  
 RESULT 7  
 AAL51569  
 ID AAL51569 standard; DNA; 2597 BP.  
 AC AAL51569;  
 XX 10-APR-2003 (first entry)  
 XX Human nucleic acid-associated protein coding sequence - SEQ ID NO 52.  
 XX Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;  
 XX cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;  
 XX cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;  
 XX mental retardation; neurological disorder; Alzheimer's disease; epilepsy;  
 XX Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;  
 XX Crohn's disease; transgenic animal; animal model.  
 XX Homo sapiens.  
 XX WO2003000864-A2.  
 XX 03-JAN-2003.  
 XX 20-JUN-2002; 2002WO-US021179.  
 XX 22-JUN-2001; 2001US-0300518P.  
 XX 29-JUN-2001; 2001US-0301787P.  
 XX 29-JUN-2001; 2001US-0301792P.  
 XX 29-JUN-2001; 2001US-0301892P.  
 XX 29-JUN-2001; 2001US-0301893P.  
 XX 06-JUL-2001; 2001US-0303405P.  
 XX 06-JUL-2001; 2001US-0303442P.  
 XX 15-MAR-2002; 2002US-0364438P.  
 XX (INCY-) INCVTE GENOMICS INC.  
 XX Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;  
 XX Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe LJ;  
 XX Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR, Lee SY;  
 XX Borowsky ML, Yao MG, Wallia NK, Bandman O, Lal PG, Becha SD, Lee SY;  
 XX Richardson TW, Elliott VS, Luo W, Tang YT, Zabatjadian Y, Lu Y;  
 XX WPI; 2003-201420/19.  
 XX P-PSDB; AA016419.  
 XX New nucleic acid-associated proteins and polynucleotides, useful for  
 XX diagnosing, treating or preventing cell proliferative (e.g. cancer),  
 XX neurological (e.g. epilepsy or Parkinson's disease), or autoimmune  
 XX disorders (e.g. AIDS).  
 XX Claim 12; Page 292-293; 312pp; English.  
 XX The invention comprises the amino acid and coding sequences of human  
 XX nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
 XX the invention are useful for diagnosing, treating or preventing disorders  
 XX associated with aberrant expression of NAAP, such as: cell proliferative  
 XX disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis

CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia  
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,  
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders  
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of  
CC the invention are useful for creating transgenic animals to model human  
CC disease. The present DNA sequence encodes a human nucleic acid-associated  
CC protein of the invention  
XX  
SQ Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

Query Match 7.7%; Score 94.6; DB 7; Length 2597;  
Best Local Similarity 55.1%; Pred. No. 3e-09;  
Matches 231; Conservative 0; Mismatches 179; Indels 9; Gaps 2;  
QY 78 TTGCACATATGAAGGGTGTGATAAAGCCCTATATCGACCATCATATTATAGACCAACATTT 137  
DB 1070 TTATAATGTGAAGATGTGGCAAGCCCTTTATGTCTCTCAACCTTACTCAACATAA 1129  
QY 138 AAGAACCCACGPAATGATCGACCGCTATATAATGTACAGTGGACGATTTGTGATAAGCAT 197  
DB 1130 GAGAAATTCATCTGGAGAGAAACCTTACAAATGT-----GAAGAGTGTGGCAAGCCCT 1183  
QY 198 TTTCAGAAATACATTTGGAACACATATGTGATCACAATTCGGAATAAAACCCATTCCA 257  
DB 1184 TAACGTGCTCTCACTTCTACTCAACATTAAGAGATTCATCTGGAGAAAACCCATACMA 1243  
QY 258 TTGTTTCAGTGTGTGTAAGGGGTTAATCTCGACAAACATTTGAAAGACATGAATCAC 317  
DB 1244 ATGTGAAGATGTGGCAAGCCCTTTAACACATCTCTCACACCTCACCACACATAAAGAA 1303  
QY 318 CCATACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCATTTTATAA---ACA 374  
DB 1304 TCATACCGGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCCTTTAACCGATTCTC 1363  
QY 375 TCAATCTTTAAGACATCAATATATCTGTTCATGAAAAAACAATTAACGTGTAAACAATG 434  
DB 1364 ACACCTTACTACACATAAGATATTCATCTGGAGAGAAACCCCTACAAATGTAAAGATG 1423  
QY 435 TAATAAGTTTTCACCTCGACCTTCAAAATTTAGCAACAATTAATTAACATCATCTGGTG 493  
DB 1424 TGGCAAGGCTTTTAAAGGGTCTCTCAAACTTACTGAACATAGGATAATTCATCTGGAG 1482

RESULT 8  
AAS64586 standard; cDNA; 3639 BP.  
XX  
AC AAS64586;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #390.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX  
XX P-PSDB; ABG00399.

XX  
PT  
PT  
PT  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity.  
Claim 1; SEQ ID NO 390; 103pp; English.  
The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
and in recombinant production of (II). The polynucleotides are also used  
in diagnostics as expressed sequence tags for identifying expressed  
genes. (I) is useful in gene therapy techniques to restore normal  
activity of (II) or to treat disease states involving (II). (II) is  
useful for generating antibodies against it, detecting or quantitating a  
polypeptide in tissue, as molecular weight markers and as a food  
supplement. (II) and its binding partners are useful in medical imaging  
of sites expressing (II). (I) and (II) are useful for treating disorders  
involving aberrant protein expression or biological activity. The  
polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
coding sequences of the invention. Note: the sequence data for this  
patent did not appear in the printed specification, but was obtained in  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

Query Match 7.5%; Score 92.4; DB 5; Length 3639;  
Best Local Similarity 54.1%; Pred. No. 8.6e-09;  
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

QY 69 AAAGTATATTGTCACATATGAAGGGTGTGATAAAGCCCTATAATCGACCATCATTTATAGA 128  
DB 738 AGAAGAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTAGCCATCTTCAACCCITGC 797  
QY 129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATTAATGTACAGTGGACGATTTGA 188  
DB 798 TAAACATAAGAGAAATTCATCTGGAGAGAAACCCCTACAAATGT-----GAAGAAATGTGG 851  
QY 189 TAAAGCATTTTTCAGAAAAATCACATTTTGGAAACACATATTGTATCATCATCCGAAAAAAA 248  
DB 852 CAAAGCTTTTAGCCGTTCTTCAACCCCTTGCTAAACATTAAGAGATTCATCTGGAGAGAA 911  
QY 249 ACCATTCATTTGTTTCAGTGTGTGTAAGGGGTTTAAATCTCGACCAACATTTGTAAGACAA 308  
DB 912 ACCCTACAAATGTAAAGAAATGTGGCAAGCTTTTAGCAATTCCTCAACCCCTTGTCTAATCA 971  
QY 309 TGAATCACCCTACCAAGTCATTTAATGTACATTTTGAATAATGTCAAGAGCATTTTA 368  
DB 972 TAAATTAATCTATCTGAGAGAAACCCCTACAAATGTAAAGATGTGACAAAGCTTTTAA 1031  
QY 369 TAAACATCAATC---TTTAAAGACATCATATATTATCTGTTTATGAAAAAACATTTAAGCTG 425  
DB 1032 CGGACTCTCAACCCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACCTTACAAATG 1091  
QY 426 TAAACATGTAATAAGTTTTCATCTGACCTTCAAAATTTAGCACCAACATATAAATAACA 485  
DB 1092 TGAAGAAATGTGGCAAGCTTTTAAATCGATCTTCAATCTTACTATACATAGTTTATCA 1151  
QY 486 TCATGGTGGATCTCCT 501  
DB 1152 TACTGGAGAGAAACCT 1167

RESULT 9  
ADB63624  
ID ADB63624 standard; cDNA; 2110 BP.  
XX

AC ADB63624;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human cDNA encoding clone THYMU20071120.  
XX  
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
KW tissue regeneration; cell regeneration; membrane protein;  
KW signal transduction-related protein; transcription-related protein;  
KW osteoporosis; neurological disease; cancer; tumour.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Location/Qualifiers  
FT 170..1792  
FT /\*tag= a  
FT /product= "Clone THYMU20071120 protein"  
XX  
FN EP1308459-A2.  
XX  
XX 07-MAY-2003.  
XX  
XX 28-MAR-2002; 2002EP-00007401.  
XX  
XX 05-NOV-2001; 2001JP-00379298.  
XX  
XX 25-JAN-2002; 2002US-00350978.  
XX  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I,  
XX Seki N, Yoshikawa T, Otsuka M, Negahari K, Masuho Y;  
XX WPI; 2003-450961/43.  
XX P-PSDB; ADB65594.  
XX  
XX New polynucleotides and polypeptides, useful for developing a diagnostic  
XX marker or medicines for regulation of their expression and activity, or  
XX PT as targets of gene therapy.  
XX  
XX Claim 1; Page; 222pp; English.  
XX  
XX The invention discloses a polynucleotide comprising a sequence selected  
XX from 1970 fully defined nucleotide sequences which encode novel  
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
XX or its partial peptide, an antibody binding to the polypeptide or peptide  
XX of the polynucleotide, immunologically assaying the polypeptide or  
XX peptide of the polynucleotide by contacting the polypeptide or peptide  
XX with the antibody of the encoded protein, and observing the binding  
XX between the two, a transformant carrying the polynucleotide in an  
XX expressible manner and an antisense polynucleotide. The oligonucleotide  
XX is useful as a primer for synthesising the polynucleotide, or as a probe  
XX for detecting the polynucleotide. The polynucleotides and encoded  
XX proteins are useful as pharmaceutical agents and many disease-related  
XX genes may be included in them, for developing a diagnostic marker or  
XX medicines for regulation of their expression and activity, or as targets  
XX of gene therapy. The genes are involved in tissue and/or cell  
XX regeneration. Membrane proteins, signal transduction-related proteins,  
XX transcription-related proteins, disease-related proteins and genes  
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,  
XX neurological diseases, cancer, tumours). The cDNA may be used to regulate  
XX the activity or expression of the encoded protein to treat diseases. The  
XX sequence presented is a cDNA of the invention. Note: Some of the sequence  
XX data for this patent is not represented in the printed specification, but  
XX is based on sequence information supplied by the European Patent Office.  
XX  
XX Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;  
SQ  
Query Match 7.4%; Score 91.8; DB 9; Length 2110;  
Best Local Similarity 51.7%; Pred. NO. 1e-08;  
Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;  
AC 61 CGTCCCAAAAAGTATATTTGACATATGAAGGGTGTGATAAGCCCTATAATCGACCATCA 120  
DB 1253 CATACCGAAGAGAAACCCCTACAAATGTAAAGATGTGGCAAGCTTTTAAACACCTCTCA 1312  
QY 121 TTATTAGAGCAACATTTTAAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180  
DB 1313 GCCCTTACTACATAAGAGAAATTCACACTGGAGAGAAACCCCTACAAATGT-----GAA 1366  
QY 181 GATTGTGATAAAGCATTTTTTTCAGAAAATACATTTTGGAAAACACATATTTGTATCATATCC 240  
DB 1367 GAATGTGGCAAGCTTTTAAACCGATCCTCAAAACCTTACTGACATAGAAGAACTTCATACT 1426  
QY 241 GAAAAAACCATTCCATTTGTTTCAGTGTGTGTGTAAAGGGTTAATTTCTCGACACACTTG 300  
DB 1427 GGAAGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTATCCAAATCCTCAAAACCT 1486  
QY 301 AAAAGACATGAATCACCACATACAAAGTCATTTAAATGTACATTTGAAAAATTTGTCAAGAA 360  
DB 1487 ACTGACATATAAATAATTCATTTCTGAGAGATACCCCTACAAAGTGTGAAGATGTGGCAAA 1546  
QY 361 GCATTTTATAACATCAATCTTTA---AGACATCATATATATCTGTTTCATGAAAAAACA 417  
DB 1547 GCTTTTAAACACTCCTCATCTTACTACATACATAAAGAAATTCATCTACTGGGAGAACCC 1606  
QY 418 TTAAGCTGTAAACAATGTAATAAAGTTTTTCACTCGACCTTCAAAATTTAGCAACAACATAA 477  
DB 1607 TACAAATGTGAAGATGTGGCAAGCTTTTAGCCGATCCTCAAAACCTTACTGACATAG 1666  
QY 478 TTAACAATCATGTGTG-ATCTCTCTGCTTTTCAATGTGATCATCTCTGTTGTTTAAAAA 536  
DB 1667 ATAATTCATCTGGAGAGAAACCCCTATAAATGTGAGAGATGTGACAAAGCTTTTAAACCA 1726  
QY 537 TTTCCAAACTTGGTCAGTATTACAAATTTTATATAAACAACATGTCATCCAAAACCTTAAATG 596  
DB 1727 TCTGCAACCT-----TACTAAACATATAAATAATACATCTGGAGAGAACTACAGAAC 1780  
QY 597 TCCTAAATGTGTAAGGTGTGTTGGGAAAAAAGGTTTATCTTCATATATTTAAGTCA 656  
DB 1781 TCGAATGTGTGATAATGATTTTGACAAACACCTTCAAAATTTTCTAAATATAAGGAAATCA 1840  
QY 657 TGATGAT 663  
DB 1841 TACTGGT 1847  
RESULT 10  
ADC56695  
ID ADC56695 standard; cDNA; 2760 BP.  
XX  
AC ADC56695;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Human macroprotein-45-76 cDNA.  
XX  
XX human; macroprotein-45.76; dementia; arrhythmia; asthma; diabetes; ss;  
XX gene.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Location/Qualifiers  
FT 180..1430  
FT /\*tag= a  
FT /product= "Human macroprotein-45.76"  
XX  
XX CN1382725-A.  
XX  
XX 04-DEC-2002.  
XX  
XX 26-APR-2001; 2001CN-00112764.  
XX  
XX 26-APR-2001; 2001CN-00112764.  
XX

(BIOW-) BLOWDOWN GENE DEV INC SHANGHAI.  
Mao Y, Xie Y;  
WPI: 2003-269486/27.  
P-PSDB; ADC56696.  
New human macroprotein-45.76, encoding polynucleotide, antagonist and recombinant production, useful for treating dementia, arrhythmia, asthma and diabetes.  
Claim 6; SEQ ID NO 1; 34pp; Chinese.  
The invention relates to a novel human macroprotein-45.76, the encoding polynucleotide, an antagonist and a method of recombinant production. The protein of the invention may be useful for treating dementia, arrhythmia, asthma and diabetes. The current sequence is that of the human macroprotein-45.76 cDNA of the invention.  
Sequence 2760 BP; 1016 A; 463 C; 499 G; 782 T; 0 U; 0 Other;  
Query Match 7.4%; Score 91.8; DB 9; Length 2760;  
Best Local Similarity 52.1%; Pred. No. 1.1e-08;  
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;  
QY 11 GTGACGAAACCAATCGATATCATCTTTAATATCTCTTCTTCTTCATCAGTCCCAAAA 70  
DB 712 GTGGCAATCATTTTGCATGCTTTCACATTAATCTCAACATAGAAAATCTACTAGAG 771  
QY 71 AGTATATTTTCACATATGAGGTGTGATGAAGCCTTATATGACCATCATTTATAGAGC 130  
DB 772 AGTATCTTACAAATGTGAAGATGTGTAAGCCCTTAACTGGTCTCAACCCCTTACTA 831  
QY 131 AACATTTAAGAACCCAGTAAATGATGCGACCGTATAATGTACAGTGGAGATTGTGATA 190  
DB 832 AACATAAGATTAATTCATCTGAGAGAAAACCCCTACAAATGT-----GAAGATGTGGCA 885  
QY 191 AAGCATTTTTCAGAAAATCATTTTGAACACATATTTGATCACAATTCGAAAAAATAAC 250  
DB 886 AGCTTTTACCGTCTCTCAATCTTACTAAACATATAATATTCATCTGGAGAGAAAC 945  
QY 251 CATTCCATTTGTCAGTGTGGTAAAGGGTTAATCTCTGACAACTTGAAGACATG 310  
DB 946 CCTACAAATGTGAAGATGTGGCAAGCTTTTAAACCGTCTCAACCCCTTACTAAACATA 1005  
QY 311 AATCACCACATACAAAGTCATTTAAATGTGATTTGAAATTTGCAAGAGCATTTTATA 370  
DB 1006 AAGAATTCATCAGAGAGAAACCCCTACAAATGTGAGATGTGCAAGGCTTTAAACC 1065  
QY 371 AACATCAATCTTT---AAGACATCATATATATCTGTTCATGAAAAAATTAACCTGTA 427  
DB 1066 AGTTCTCGATTCTTATAAACAATAAGAGAAATTCATATGGAAGATAAACCCCTACAAATGTG 1125  
QY 428 AAGAATGTAATAAGTTTTCACCTGACCTTCAAAATTTAGCACAACATAAATTAACATC 487  
DB 1126 AAGAATGTGCAAGCCCTTTAGAGTATCTCAATTTCTTAAACATATAGATATCCATA 1185  
QY 488 ATGTTGGATCTCC 500  
DB 1186 CTGGGGAATAAAC 1198  
RESULT 11  
ADB62468  
ID ADB62468 standard; cDNA; 3078 BP.  
AC ADB62468;  
XX  
XX  
XX 04-DEC-2003 (first entry)  
XX Human cDNA encoding clone FEBRA20063720.  
XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
KW tissue regeneration; cell regeneration; membrane protein; signal transduction-related protein; transcription-related protein; osteoporosis; neurological disease; cancer; tumour.  
XX Homo sapiens.  
XX Location/Qualifiers  
XX Key 118..1878  
XX CDS /\*tag= a  
XX /product= "Clone FEBRA20063720 protein"  
XX  
XX EPI308459-A2.  
XX  
XX 07-MAY-2003.  
XX 28-MAR-2002; 2002EP-00007401.  
XX 05-NOV-2001; 2001JP-00379298.  
XX 25-JAN-2002; 2002US-00350978.  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX WPI: 2003-450961/43.  
XX P-PSDB; ADB64438.  
XX New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.  
XX Claim 1; Page; 222pp; English.  
XX The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesizing the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.  
XX Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;  
Query Match 7.4%; Score 91.8; DB 9; Length 3078;  
Best Local Similarity 52.1%; Pred. No. 1.1e-08;  
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;  
QY 11 GTGACGAAACCAATCGATATCATCTTTAATATCTCTTCTTCTTCATCAGTCCCAAAA 70  
DB 650 GTGGCAATCATTTTGCATGCTTTCACAAATTAACATCAAGAAAATTCATCTAGAG 709  
QY 71 AGTATATTTGCACATATGAGGTGTGATGAAGCCTATAATCGACCATCATTTATAGAGC 130  
DB 710 AGTATTTTCAAAATGTGAAGATGTGTAAGCCCTTAACTGGTCTCTCAACCCCTTACTA 769

PA (BIOW-) BLOWDOWN GENE DEV INC SHANGHAI.  
Mao Y, Xie Y;  
WPI: 2003-269486/27.  
P-PSDB; ADC56696.  
New human macroprotein-45.76, encoding polynucleotide, antagonist and recombinant production, useful for treating dementia, arrhythmia, asthma and diabetes.  
Claim 6; SEQ ID NO 1; 34pp; Chinese.  
The invention relates to a novel human macroprotein-45.76, the encoding polynucleotide, an antagonist and a method of recombinant production. The protein of the invention may be useful for treating dementia, arrhythmia, asthma and diabetes. The current sequence is that of the human macroprotein-45.76 cDNA of the invention.  
Sequence 2760 BP; 1016 A; 463 C; 499 G; 782 T; 0 U; 0 Other;  
Query Match 7.4%; Score 91.8; DB 9; Length 2760;  
Best Local Similarity 52.1%; Pred. No. 1.1e-08;  
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;  
QY 11 GTGACGAAACCAATCGATATCATCTTTAATATCTCTTCTTCTTCATCAGTCCCAAAA 70  
DB 712 GTGGCAATCATTTTGCATGCTTTCACATTAATCTCAACATAGAAAATCTACTAGAG 771  
QY 71 AGTATATTTTCACATATGAGGTGTGATGAAGCCTTATATGACCATCATTTATAGAGC 130  
DB 772 AGTATCTTACAAATGTGAAGATGTGTAAGCCCTTAACTGGTCTCAACCCCTTACTA 831  
QY 131 AACATTTAAGAACCCAGTAAATGATGCGACCGTATAATGTACAGTGGAGATTGTGATA 190  
DB 832 AACATAAGATTAATTCATCTGAGAGAAAACCCCTACAAATGT-----GAAGATGTGGCA 885  
QY 191 AAGCATTTTTCAGAAAATCATTTTGAACACATATTTGATCACAATTCGAAAAAATAAC 250  
DB 886 AGCTTTTACCGTCTCTCAATCTTACTAAACATATAATATTCATCTGGAGAGAAAC 945  
QY 251 CATTCCATTTGTCAGTGTGGTAAAGGGTTAATCTCTGACAACTTGAAGACATG 310  
DB 946 CCTACAAATGTGAAGATGTGGCAAGCTTTTAAACCGTCTCAACCCCTTACTAAACATA 1005  
QY 311 AATCACCACATACAAAGTCATTTAAATGTGATTTGAAATTTGCAAGAGCATTTTATA 370  
DB 1006 AAGAATTCATCAGAGAGAAACCCCTACAAATGTGAGATGTGCAAGGCTTTAAACC 1065  
QY 371 AACATCAATCTTT---AAGACATCATATATATCTGTTCATGAAAAAATTAACCTGTA 427  
DB 1066 AGTTCTCGATTCTTATAAACAATAAGAGAAATTCATATGGAAGATAAACCCCTACAAATGTG 1125  
QY 428 AAGAATGTAATAAGTTTTCACCTGACCTTCAAAATTTAGCACAACATAAATTAACATC 487  
DB 1126 AAGAATGTGCAAGCCCTTTAGAGTATCTCAATTTCTTAAACATATAGATATCCATA 1185  
QY 488 ATGTTGGATCTCC 500  
DB 1186 CTGGGGAATAAAC 1198  
RESULT 11  
ADB62468  
ID ADB62468 standard; cDNA; 3078 BP.  
AC ADB62468;  
XX  
XX  
XX 04-DEC-2003 (first entry)  
XX Human cDNA encoding clone FEBRA20063720.  
XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
KW tissue regeneration; cell regeneration; membrane protein; signal transduction-related protein; transcription-related protein; osteoporosis; neurological disease; cancer; tumour.  
XX Homo sapiens.  
XX Location/Qualifiers  
XX Key 118..1878  
XX CDS /\*tag= a  
XX /product= "Clone FEBRA20063720 protein"  
XX  
XX EPI308459-A2.  
XX  
XX 07-MAY-2003.  
XX 28-MAR-2002; 2002EP-00007401.  
XX 05-NOV-2001; 2001JP-00379298.  
XX 25-JAN-2002; 2002US-00350978.  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX WPI: 2003-450961/43.  
XX P-PSDB; ADB64438.  
XX New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.  
XX Claim 1; Page; 222pp; English.  
XX The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesizing the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.  
XX Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;  
Query Match 7.4%; Score 91.8; DB 9; Length 3078;  
Best Local Similarity 52.1%; Pred. No. 1.1e-08;  
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;  
QY 11 GTGACGAAACCAATCGATATCATCTTTAATATCTCTTCTTCTTCATCAGTCCCAAAA 70  
DB 650 GTGGCAATCATTTTGCATGCTTTCACAAATTAACATCAAGAAAATTCATCTAGAG 709  
QY 71 AGTATATTTGCACATATGAGGTGTGATGAAGCCTATAATCGACCATCATTTATAGAGC 130  
DB 710 AGTATTTTCAAAATGTGAAGATGTGTAAGCCCTTAACTGGTCTCTCAACCCCTTACTA 769

QY 131 AACATTAAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGACGAGTGTGATA 190  
DB |||||  
QY 770 AACATAAGATAATTCATCTACTGAGAGAAACCCCTACAAATGT-----GAAGAAATGTGGCA 823  
DB |||||  
QY 191 AAGCATTTTTCAGAAAAACACATTTTGGGAAACACATATTTGTATCATCAATTCGAGAAAAAAC 250  
DB |||||  
QY 824 AAGCTTTTAAACCGGTCTCAATCTTACTAAACATAAAAAATAATTCATCTCGAGAGAAAC 883  
DB |||||  
QY 251 CATTCATTTTTCAGTGTGTGTAAGGGTTAATTTCTCGACACACATTTGAAAAAGACATG 310  
DB |||||  
QY 884 CTTCAAAATGTGAAGATGTGGCAAGGCTTTTACCGGTCTCTCAACCTTACTAAACATA 943  
DB |||||  
QY 311 AAATCACCACCATCAAAAGTCATTTAAATGTACATTTTGAATAATTTGCAAGACATTTTATA 370  
DB |||||  
QY 944 AAAGAATTCATACAGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGGCTTTTAAAC 1003  
DB |||||  
QY 371 ACATCAATCTTT---AGACATCATATATATCTGTTCATGAAAAACATTAACGTGTA 427  
DB |||||  
QY 1004 AGTCTCGAATCTTAAATAACATAGAGAAATTCATATGAGATTAACCCCTACAAATGTG 1063  
DB |||||  
QY 428 AACAATGTAATAAGTTTTCATCGACCTTCAAAATTTAGCAACATATAAATTAACAATC 487  
DB |||||  
QY 1064 AAGAATGTGGCAAGGCTTTAGAGTATTTCAATTTCTTAAAAACATTAAGATAATCCATA 1123  
DB |||||  
QY 488 ATGTTGGATCTCC 500  
DB |||||  
QY 1124 CTGGGGAATAAAC 1136  
DB |||||

RESULT 12  
AAD55863  
ID AAD55863 standard; cDNA; 2026 BP.  
XX  
AC AAD55863;  
XX  
DT 07-AUG-2003 (first entry)  
XX  
DE Human nucleic acid associated protein (NAAP)-34 cDNA.  
XX  
KW Human; nucleic acid associated protein; NAAP; stroke; AIDS; neotropic;  
cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;  
developmental disorder; antiinflammatory; neuroprotective; thyromimetic;  
Cushing's syndrome; infection; gene therapy; cytostatic; anticonvulsant;  
cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;  
gene, ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT 126..1973  
FT CDS  
FT /\*tag= a  
FT /product= "Human NAAP protein"  
XX  
PN WC2003006618-A2.  
XX  
PD 23-JAN-2003.  
XX  
PE 10-JUL-2002; 2002WO-US021971.  
XX  
PR 12-JUL-2001; 2001US-0305089P.  
PR 12-JUL-2001; 2001US-0305104P.  
PR 13-JUL-2001; 2001US-0305325P.  
PR 13-JUL-2001; 2001US-0305390P.  
PR 19-JUL-2001; 2001US-0306960P.  
PR 20-JUL-2001; 2001US-0306694P.  
PR 27-JUL-2001; 2001US-0308170P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;  
PI Yue H, Baughn MR, Emerling BM, Lal PG, Lu DM, Forsythe IJ;  
PI Ramkumar J, Li JX, Becha SD, Duggan BM, Sanjanwala MM, Lee EA;

PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yac MG;  
PI Barroso I, Tran B, Walia NK, Hafalia AJA, Nguyen DB, Lu Y;  
PI Arvizu CS;  
XX  
DR WPI: 2003-221732/21.  
DR P-PSDB; AAE37047.  
XX  
PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
PT treating and preventing diseases or conditions associated with the  
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
PT infections.  
XX  
PS Claim 5; Page 259-260; 260pp; English.  
XX  
CC The invention relates to human nucleic acid associated proteins (NAAP)  
CC and their corresponding nucleic acid sequences. The invention is useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression or overexpression of NAAP, such as cell  
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
CC exogenous compounds on the expression of nucleic acid and amino acid  
CC disorders, or infections. It is also useful in assessing the effects of  
CC compounds for which acts as their agonist or antagonist. The microarray  
CC is useful in monitoring or measuring protein-protein interactions, drug-  
CC target interactions, and gene expression profiles. NAAP DNA is used in  
CC gene therapy. The present sequence is human NAAP cDNA  
XX  
SQ Sequence 2026 BP; 699 A; 422 C; 396 G; 509 T; 0 U; 0 Other;

Query Match 7.3%; Score 90.4; DB 7; Length 2026;  
Best Local Similarity 51.8%; Pred. No. 2e-08;  
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;  
QY 7 GAAAGTCAGCAACCAAAATCGATATCATCTTTAAATATCTTCTTCTTCTCATCAGCTCCC 66  
DB 1002 GAATGTGGCAAGGCTTCTACCATTTTACACCTTACTACATCAAGTAATTCATCTACT 1061  
QY 67 AAAAAATATATTTTCACATATGAAGGGTGTGATAAAGCCCTATAATCGACATCATTTATA 126  
DB 1062 GGAGAGAGGCCCTTCAAAATGTGAAGATGTGGTAAAGCTTTTAAACACCCCTTCAGCCCT 1121  
QY 127 GAGCAATTTAAGAACCCACAGTATGATCGACCGTATAAATGTACAGTGGACGATTGT 186  
DB 1122 ACTACATAGTTCATTCATGTGTTAAAGAAAAACCCCTACAAATGT-----GAAGATGT 1175  
QY 187 GATAAAGCATTTTTCAGAAATCACATTTTGGAAACACATATTTGATCATCTCCGAAAAA 246  
DB 1176 GACAAAGCTTTTAAACGATTCATACCTTACTTAACATAAGATAATTCATTTCTGGAGAG 1235  
QY 247 AAACCATTCATTTTCAGTGTGTGTAAGGGGTTAATTTCTCGACACACATTTGAAAGA 306  
DB 1236 AAATCTTACAAATGTGAACAAATGTGCAAGGCTTTTAACTGGTCTTCAACCCCTTACAAA 1295  
QY 307 CATGAATCACCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAAAGCATTT 366  
DB 1296 CATAGAAGATTCATCTGGAGAGAACCCCTACAAATGTGAAGATGTGCAAGGCTTT 1355  
QY 367 TATAAACAT---CAATCTTTAAGACATCATATATTTCTTTCATGAAAAACATTAACG 423  
DB 1356 AATGTGCTTTTACACCTTACTACATAGATGATTCATCTGGAGAGAAACCCCTACAAA 1415  
QY 424 TGTAACAATGTAAATTAAGTTTTCATCGACCTTCAAAATTTAGCAACACATAAATTAATA 483  
DB 1416 TGTGAAGATGTGGCAAGGCTTTTAACTCTCTCAAACTTACTATACATAGATAATT 1475  
QY 484 CATCATGGTGGATCTCTCT 501  
DB 1476 CATACTGGAGAGAAACCT 1493

RESULT 13

ADAS3124	DB	1318	CATAGAGAAATTCATCTGGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCTTT	1377
ID ADAS3124 standard; cDNA; 2114 BP.				
XX AC				
ADAS3124;				
XX DT				
20-NOV-2003 (first entry)				
XX DE				
Human coding sequence, SEQ ID 692.				
XX KW				
Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;				
XX KW				
Gene Therapy; human; secretory protein; membrane proteins; cancer;				
XX KW				
inflammatory disease; osteoporosis; neurological disease; gene; ss.				
XX OS				
Homo sapiens.				
XX XX				
EP1293569-A2.				
XX PD				
19-MAR-2003.				
XX PF				
21-MAR-2002; 2002EP-00006586.				
XX PR				
14-SEP-2001; 2001JP-00328381.				
XX PR				
24-JAN-2002; 2002US-0350435P.				
XX PA				
(HELI-) HELIX RES INST.				
XX PA				
(SEAS-) RES ASSOC BIOTECHNOLOGY.				
XX PI				
Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;				
XX PI				
Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;				
XX PI				
Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;				
XX XX				
WPI; 2003-395539/38.				
XX DR				
P-PSDB; ADA54763.				
XX XX				
New polynucleotides encoding full-length polypeptides, e.g. secretory				
XX PT				
and/or membrane proteins, useful for developing medicines for diseases in				
XX PT				
which the gene is involved, or as target molecules for gene therapy.				
XX PS				
Claim 1; SEQ ID NO 692; 205pp; English.				
XX CC				
The present invention relates to novel human secretory or membrane				
XX CC				
proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-				
XX CC				
ADA54071). The coding sequences are useful in the gene therapy of				
XX CC				
diseases caused by abnormalities of the proteins, e.g. cancer,				
XX CC				
inflammatory diseases, osteoporosis or neurological disease.				
XX XX				
Sequence 2114 BP; 717 A; 439 C; 434 G; 524 T; 0 U; 0 Other;				
Query Match				
Best Local Similarity				
Matches				
258; Conservative				
0; Mismatches				
231; Indels				
9; Gaps				
2;				
QY				
7				
GAAAGTGACGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCAGTCCC				
66				
DB				
1024				
GAAATGTGGCAAGCCTTCTACCATCTTTCACACCTTACTACACATAGGTAATTCATACT				
1083				
QY				
67				
AAAAGTATATTCACATATGAAGGTGTGATAAAGCCTATATCGACCATCATTTATTA				
126				
DB				
1084				
GGAGAGAGCCCTTCAATGTGAAGATGTGGTAAAGCTTTTACACCCCTTCAGCCCTT				
1143				
QY				
127				
GAGCAACATTTAAGAACCCACAGTAATGATCGACCGGTATAAATGTACAGTGGAGATTGT				
186				
DB				
1144				
ACTACACATAGTTCATTCATGTGTTAAAGAAACCCCTACAAATGT-----GAAGATGT				
1197				
QY				
187				
GATAAGCATTTTTCAGNAATTCATTTGAAACACATATTTGATCAGATTCGGAATAA				
246				
DB				
1198				
GACAAAGCTTTTAAACCGATTCTTACCTTACTAAACATAGATAANTTCATTCGGAGAG				
1257				
QY				
247				
AAACCATTCCTATTTTCAAGTGTGGTAAAGGGGTAAATTCGACACACATTTGAAAGA				
306				
DB				
1258				
AAATCTTACAATGTGACATATGGCGAAGCTTTTAACTGGTCTTCAACCCCTTACAAA				
1317				
QY				
307				
CATGAATCATCCCATACAAAGTCATTTAATGTACATTTGAAATTTGTCAAGAGCATTT				
366				

DB	1318	CATAGAGAAATTCATCTGGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCTTT	1377
QY	367	TATAAACAT---CAATCTTTAAGACATCATATATCTTCTTCATGAAACCAATTAAACG	423
DB	1378	AATGTGCTTTCACACCTTACTACACATAGATGATTCATCTGGAGAGAAACCCCTACAAA	1437
QY	424	TGTAAACAATGTAATAAAGTTTTCACCTCGACCTTCAAAAATTTAGCACACATAAATTTAAA	483
DB	1438	TGTGAAGAATGTGGCAAGCCTTTAACCACTCTCTCAAAACCTTACTATACATAGATAATT	1497
QY	484	CATCATGGTGGATCTCTT 501	
DB	1498	CATCTGGAGAGAAACCT 1515	
RESULT 14			
ADAS2931			
ID ADAS2931 standard; cDNA; 2230 BP.			
XX AC			
ADA52931;			
XX DT			
20-NOV-2003 (first entry)			
XX DE			
Human coding sequence, SEQ ID 499.			
XX KW			
Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;			
XX KW			
Gene Therapy; human; secretory protein; membrane proteins; cancer;			
XX KW			
inflammatory disease; osteoporosis; neurological disease; gene; ss.			
XX OS			
Homo sapiens.			
XX XX			
EP1293569-A2.			
XX PD			
19-MAR-2003.			
XX PF			
21-MAR-2002; 2002EP-00006586.			
XX PR			
14-SEP-2001; 2001JP-00328381.			
XX PR			
24-JAN-2002; 2002US-0350435P.			
XX PA			
(HELI-) HELIX RES INST.			
XX PA			
(SEAS-) RES ASSOC BIOTECHNOLOGY.			
XX PI			
Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;			
XX PI			
Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;			
XX PI			
Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;			
XX XX			
WPI; 2003-395539/38.			
XX DR			
P-PSDB; ADA54570.			
XX XX			
New polynucleotides encoding full-length polypeptides, e.g. secretory			
XX PT			
and/or membrane proteins, useful for developing medicines for diseases in			
XX PT			
which the gene is involved, or as target molecules for gene therapy.			
XX PS			
Claim 1; SEQ ID NO 499; 205pp; English.			
XX CC			
The present invention relates to novel human secretory or membrane			
XX CC			
proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-			
XX CC			
ADA54071). The coding sequences are useful in the gene therapy of			
XX CC			
diseases caused by abnormalities of the proteins, e.g. cancer,			
XX CC			
inflammatory diseases, osteoporosis or neurological disease.			
XX XX			
Sequence 2230 BP; 831 A; 434 C; 400 G; 565 T; 0 U; 0 Other;			
Query Match			
Best Local Similarity			
Matches			
332; Conservative			
0; Mismatches			
313; Indels			
14; Gaps			
4;			
QY			
7			
GAAAGTGACGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCAGTCCC			
66			
DB			
656			
GAAATGTGCAAAATCACTTTGCATGTCTTTCACGCTTCACTCAACATAAAAAATTCATACT			
715			

xx	sq	Sequence	2054 BP;	703 A;	375 C;	449 G;	537 T;	0 U;	0 Other;
		Query Match	7.2%;	Score	89.2;	DB	9;	Length	2064;
		Best Local Similarity	53.7%;	Pred. No.	3.4e-08;				
		Matches	234;	Conservative	0;	Mismatches	193;	Indels	9;
									Gaps
Qy	69	AAAGTATATTGACACATATGAAGGGTGTGATAAAGCCTATATAATCGACCATCATTTAGTA	128						
Db	828	AGAGAAACCTTACAAATGTGAAGAAATGTGCACAAAGTTTTTTAGTCGCTTAATCACACCTTGA	887						
Qy	129	GCACATTTAAGAACCCACAGTAAATGATCACCGGTATAAATGTATGACGTGACGAGTGTGA	188						
Db	888	AAGACATAGAGAAATTCATCTCGAGAGAAACCGTACAAATGTAAGGT-----TTGTGA	941						
Qy	189	TAAAGCATTTTTTCAGAAAATTCATTTGGAAAACACATATTTGTATCATTTCCGAAAAAAA	248						
Db	942	CAAGGCTTTCAGACGTGATTCACACCTGGCACAACATATTTGTAATTCACACTGGAGAA	1001						
Qy	249	ACCATTCATTTGTTTCAGTGTGTGGTAAGGGGTTAAATTCGACAAACACTTGAAGAACACA	308						
Db	1002	ACCTTACAAGTGTAAATGAGTGTGCGAAGACCTTTGTTCAAAATTCATCTCTTGTGAATGCA	1061						
Qy	309	TGAAATCAACCCATAC----AAAGTCATTTAAATGTACATTTTGAAATTTGTCAAGAAGCAAT	365						
Db	1062	TAGGTCATTCATCTCGAGAGAAATGTTTCAAGTGTAAATGTAATGTGGCAAGGTTTTTAA	1121						
Qy	266	TTATATACATCCATCTCTTAAACATCATATATATATCTGTTTCATGAAAAAACATTAACGTG	425						

Db	1122	TCACAAATCAAACCTTGCATGTCATCATAGACTTCATCTGGAGAGAAACCTTACAAGTG	1181
Qy	426	TAAACCAATGTAATAAAGTTTTCACCTCGACCTTCAAAATTAGCACACATAAAATTAAAAACA	485
Db	1182	TAATGATGTGGCARGGTTTTTAATTGAAAAATCAAAACCTTGAACATCATATAGAGTTCA	1241
Qy	486	TCATGGTGGATCTCCT	501
Db	1242	TATCGAGAGAAAACCT	1257

Search completed: May 9, 2004, 06:07:03  
Job time : 557.339 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 05:31:45 ; Search time 3610.62 Seconds  
(without alignments)  
10222.531 Million cell updates/sec

Title: US-09-831-804-1\_COPY\_720\_1955

Perfect score: 1236

Sequence: 1 atgagtgaagtgaagaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hic:*
9:	gb_esti:*
10:	gb_est2:*
11:	gb_hic:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	em_gss_hum:*
18:	em_gss_inv:*
19:	em_gss_pln:*
20:	em_gss_vrt:*
21:	em_gss_fun:*
22:	em_gss_man:*
23:	em_gss_mus:*
24:	em_gss_pro:*
25:	em_gss_rod:*
26:	em_gss_phg:*
27:	em_gss_vri:*
28:	gb_gssI:*
29:	gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	182.2	14.7	628	BZ298000	BZ298000 CG3957.f1
C 2	167.6	13.6	959	29	AL405671 T7 end of
C 3	98.2	7.9	1609	11	AK032220 Mus muscu
4	97.8	7.9	2647	11	BC047646 Homo sapi

5	97.6	7.9	745	9	AU123448
6	92.4	7.5	2672	11	BC036394
7	92	7.4	591	28	AQ005136
8	91	7.4	593	28	B99387
9	91	7.4	1163	13	BX414627
10	90.4	7.3	2694	11	BC022527
11	90.4	7.3	2698	11	BC037782
12	90.2	7.3	2253	11	BC028252
13	89.2	7.2	604	10	BE161630
14	89.2	7.2	3609	11	BC032590
15	88.6	7.2	535	9	AL705393
16	88.6	7.2	2174	11	AK033001
17	88.6	7.2	2330	11	BC037426
18	88.6	7.2	2622	11	AK033958
19	88.4	7.2	1084	13	BX456765
20	88	7.1	895	13	BUI195416
21	87.6	7.1	832	13	BX437291
22	87.6	7.1	1201	13	BX355654
23	87.4	7.1	1200	13	BX437758
24	87.2	7.1	579	28	AQ347265
25	87.2	7.1	687	28	AQ389266
26	87.2	7.1	907	13	BQ423752
27	87	7.0	660	14	CD770140
28	87	7.0	842	13	BU507408
29	86.8	7.0	521	28	AQ475626
30	86.8	7.0	705	28	AQ194282
31	86.8	7.0	860	28	AQ749175
32	86.8	7.0	897	13	BQ431141
33	86.8	7.0	1906	11	AK013043
34	86.6	7.0	694	12	BG696505
35	86.4	7.0	874	13	BUI171453
36	86.4	7.0	2791	11	BC020045
37	86.2	7.0	523	14	CA874049
38	86	7.0	557	10	BF817669
39	86	7.0	667	14	CF135920
40	86	7.0	685	29	AG149704
41	86	7.0	801	14	CB657347
42	85.6	6.9	711	14	CB154257
43	85.6	6.9	712	9	AI792344
44	85.4	6.9	682	12	BG818171
45	85.4	6.9	2106	11	AK030776

#### ALIGNMENTS

RESULT 1  
BZ298000/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BZ298000 628 bp DNA linear GSS 31-OCT-2002  
CG3957.f1 Candida glabrata Random Genomic Library  
Genomic clone CG3957, genomic survey sequence.

BZ298000 GI:24440936

GSS

Candida glabrata

Candida glabrata

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

1 (bases 1 to 628)

Wong, S., Farees, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.

Evidence from comparative genomics for a complete sexual cycle in

the 'asexual' pathogenic yeast Candida glabrata

Genome Biol. 4 (2), R10 (2003)

22508158

12620120

Contact: Wong S

Department of Genetics, Smurfit Institute

Trinity College Dublin

Dublin 2, Ireland

Tel: 353 1 6828319

Fax: 353 1 6798558

Email: swong@tcd.ie

Class: plasmid ends.

```

FEATURES
  source
    Location/Qualifiers
      1..628
        /organism="Candida glabrata"
        /mol_type="genomic DNA"
        /strain="CBS 138"
        /db_xref="taxon:5478"
        /clone="CG3957"
        /clone_lib="Candida glabrata Random Genomic Library"

ORIGIN
  Query Match      14.7%; Score 182.2; DB 28; Length 628;
  Best Local Similarity 59.5%; Pred. No. 3.2e-24;
  Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

QY 84 ATATGAAGGGTGTGATAAGCGCTATATCGACCATCATTTAGACGACATTTTAAAGAAC 143
DB 625 ATACGATAATTTGTGACAAAGGCGCTTTACAAAGCGCGCTTGTCTCACAGAACACCAAGATAC 566
QY 144 CCACAGTAATGATCCACCGGTATAATGTACAGTGCAGCATTTGTGATAAGCATTTTTCAG 203
DB 565 ---AGTGCATCTTGGAGGAACCTTGGAAATGTAATCAATGTAAGTTTCATTTACTAA 509
QY 204 AAATCAATTTGGAAAAACATATTTGTATACATTTCCGAAAAAACCATTCCATTGTTTC 263
DB 508 AAAGATCCACTTAGAGACACATTATACACACACAGATGAAGACCGTTTATTGTTTC 449
QY 264 AGTGTGTGTAAAGGGGTTAATTTCTCGACACACACTTGAAGAGACATGAATCAACCCATAC 323
DB 448 ATTTTGTGGAAAGGGGCTAATTTACTAGGCACACACTGAAAGCATGAGTAACTCACAC 389
QY 324 AAAGTCATTAAATGTACATTTGAAATTTGCAAGACCATTTTATAAATCAATCTTT 383
DB 388 CAAATCGTTCAATTTGTAATGAAGTTGTAATGAGAGTTTCTACAGCACCCCTCAAT 329
QY 384 AGACATCATATATTATCTGTTTCATGAAAAACATTTACGTTGAACCAATGTAATAAGT 443
DB 328 AAGGCTCATATTTTAGCAGTTTATTACAAGTCTTAAATGTCAAGATCAAGATCAACAAAG 269
QY 444 TTTCACTCGACCTTCAAAATTTAGCAACAATAAATTTAAACATCAT-----GGTGGATC 497
DB 268 TTTTCAAGACCTTACAGACTCAAAATACATAGCAGTAAATGCCCTATATSCAGTAAACCCCTG 89
QY 498 TCTGCTTATCAATGTATCATCTCTGTTGTTTAAATTTTCCAACTTGGTCAGTATT 557
DB 208 CAATGCTTATCAATGTACCTTCAGTGTGTGTTTGAAGAGTTTCAAAACATGGTCTGCGTT 149
QY 558 ACAATTTTCATATAAAACAACTGCATCCAAAATTTAAATGTCCTAAATGTGTAAGGTTG 617
DB 148 AAGATACATGTTAAATATGATCATCCGAGTTAAATGCCCTATATSCAGTAAACCCCTG 89
QY 618 TTTTGGGAAAAAGGTTTATCTTCAATATGTTTAAGTCATGATGA 662
DB 88 TGTGGGGAAGACGGTTTAAATATGCATGATGAATAATCCACGATGA 44

RESULT 2
CNS06MMP/c      959 bp      DNA      linear      GSS 17-JUN-2001
LOCUS
DEFINITION
  T7 end of clone AU00A005F10 of library AU00A from strain CBS 3082
  of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION
  AL405671
VERSION
  AL405671.1 GI:12168715
KEYWORDS
  GSS
SOURCE
  Saccharomyces kluyveri
  ORGANISM
    Saccharomyces kluyveri
    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
    Saccharomycetales; Saccharomycetaceae; Saccharomyces.
  1 (bases 1 to 959)
REFERENCE
  Souciet J.L., Aigle, M., Artiguenave, F., Blandin, G.,
  Boivin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,
  de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
  Malpertuy, A., Neuveglise, C., Olier-Kalogeropoulos, O., Poirier, S.,
  Saurin, W., Tekala, F., Toffiano-Nioche, C., Wesolowski-Louvet, M.,
  Wincker, P. and Weissenbach, J.
  Genomic exploration of the hemiascomycetous yeasts: 1. A set of
  yeast species for molecular evolution studies
  FEBS Lett. 487 (1), 3-12 (2000)
  20584711
  PUBMED
  11152876
  2 (bases 1 to 959)
  Neuveglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F.,
  Gaillardin, C. and Casaregola, S.
  Genomic exploration of the hemiascomycetous yeasts: 9.
  Saccharomyces kluyveri
  FEBS Lett. 487 (1), 56-60 (2000)
  20584719
  PUBMED
  11152884
  3 (bases 1 to 959)
  Direct Submission
  Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
  2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
  seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
  This GSS is part of a random genomic sequencing program of thirteen
  yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
  exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
  Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
  lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
  angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
  Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
  5 kb were prepared and both extremities were sequenced. See
  keywords for description of this sequence and for the sequence of
  the other extremity of this insert.
  Location/Qualifiers
    1..959
      /organism="Saccharomyces kluyveri"
      /mol_type="genomic DNA"
      /strain="CBS 3082"
      /db_xref="taxon:4934"
      /clone="AU00A005F10"
      /clone_lib="AU00A"
      /note="end : T7"
      /complement(<71..958)
      /notes="similar to Saccharomyces cerevisiae ORF YPR186c [
      P2F1 ; Trf1A (transcription initiation factor) ]"
      /evidence=not_experimental

misc_feature
  Query Match      13.6%; Score 167.6; DB 29; Length 959;
  Best Local Similarity 57.2%; Pred. No. 1.7e-21;
  Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;

ORIGIN
  QY 211 CATTTCGAAACACATATTGTATCCATTCGAAAAAACCATTCCATTGTTTCAGTGTGT 270
  DB 958 CACTTGGAAAGACATATGTTTCAGCATTCGGATATTAACCATTTTCATTGTCCTATTGT 899
  QY 271 GGTAAAGGGGTTAATTTCTCGAACACACTTTGAAAAAGACATGAATACCCATCAAGTCA 330
  DB 898 GGTAAAGGGTGTACTACAAAGCAGCAGTATAAAAGACATGAGATCACATACCAAAATCG 839
  QY 331 TTTTAAATGTACATTTGAAAATTTCTCAAGAGCATTTTATAAATCATCAATCTTTAAGACAT 390
  DB 838 TTTCAATGTGAATATGAGGTCGACAGATGTTTTTACAGCATTTCTCAACTAAGTCA 779
  QY 391 CATATATTATCTTTTCATGAAAAAACCATTAAACGTGTAAACATGTAATTAAGTTTCACT 450
  DB 778 CATACACTTTCGGTACACTTGCAGAAATTCACCTGTGAGCATTTGCGGTAAATGTTTCAA 719
  QY 451 CGACCTTCAAAATTAGCAGAACATAAATTAAGAACATCATGGTGGATCTCTCTGCT----- 504
  DB 718 AGGCCATATCGTTCGAAAACCAATCTTGCAGACATCAATGTGATGTTTGAACAA 659
  QY 505 TATCAATGTGATCATCTCTGTTGTTTTTAAATAATTTCCAACTTTGGTCAGTATTACAAATTT 564
  DB 658 TATCAATGTACTTACACGGGTGCACAGAGACTTTTAAACCTGGAGTCTGCTCTTCAGCAG 599

```

QY 565 CATATAAACAACCTGCATCCAAACTTAATATGCTCTTAATGCTGGTAAAGTTGTGTGGG 624  
 Db 598 CATATCAAGCGGACCATCCCAAGTTGCATGTAAAGTATGCTGGCAAGCTTGTGTGTGT 539  
 QY 625 AAAAAAGGTTTATCTTCCATATATGTTAAAGTCATGATCTCTACCATGATCAAAATATGG 684  
 Db 538 GAGTCGTGTACAGATGCACATGCGAGTCCATGACGAGCTTTAGTATATAAAATATGG 479  
 QY 685 ACTTGTGATTATGTTGATGTTGGGAATTTGCAAGAAATGAATAGTTGAACATTAT 744  
 Db 478 AAATGTACCAATTTGTGACGAAGTCTCTTTTGTCTAAGAAGCTGATCTTTATCTCACTAC 419  
 QY 745 AATATCTTCCATGATGTTATATATCC 770  
 Db 418 ATGACATCATTAAGGATGATATACC 393

RESULT 3  
 AK032220  
 LOCUS  
 DEFINITION  
 Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
 enriched library, clone:6430503015 product:MSZF33 (FRAGMENT)  
 homolog [Mus musculus], full insert sequence.

ACCESSION  
 AK032220  
 VERSION  
 AK032220.1 GI:26328050  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus

REFERENCE  
 1  
 Carninci, P., and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 9275253  
 MEDLINE  
 PUBMED  
 10349636

REFERENCE  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 MEDLINE  
 PUBMED  
 11042159

REFERENCE  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 MEDLINE  
 PUBMED  
 11076861

TITLE  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)

REFERENCE  
 4  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

TITLE  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216]  
 COMMENT  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/.  
 Location/Qualifiers  
 1..1609  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM\_DB:6430503015"  
 /db\_xref="MG1:2395676"  
 /db\_xref="taxon:10090"  
 /clone="6430503015"  
 /sex="male"  
 /tissue\_type="olfactory brain"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 330..1223  
 /note="unnamed protein product; MSZF33 (FRAGMENT) homolog  
 [Mus musculus] (SPTRI088251, evidence: FASTA, 87.1%ID,  
 100%length, match=255)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC27766.1"  
 /db\_xref="GI:26328051"  
 /translation="MLETYRLNNAVGYNWEDSNIEBHCESRRHGRHNRHTGKPYE  
 GIOYGEAFVHSSLSQMRKIHTGKPYKCNQCDKAYSRHSILQIKHTSGKPYECN  
 QCGAFTQSHSLKTHMVTHTGKPYKCNQCGKAFACNKLKHRTHTGKPYKDDQC  
 NKAFTVSYLVQVHKHTHTGKPYKCNQCGKAFARSHLVKHKIHTGKPYKCNQCGK  
 ALAHSILQVHQRTHTGKPYECCEQCGKAFANQSYFQVHKIHTGKPYKDCQCGKAF  
 VGSDDLKRRHVRHIGRETLQM"  
 1594..1599  
 /note="putative"  
 1609  
 /note="putative"  
 polyA\_signal  
 polyA\_site  
 ORIGIN  
 Query Match 7.9%; Score 98.2; DB 11; Length 1609;  
 Best Local Similarity 51.2%; Pred. No. 1.5e-08;  
 Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;  
 QY 60 ACGTCCCAAAAGTATATTTGCACATATGAAGGGTGTGATAAGCCTATATCGACCATC 119  
 Db 521 ACATCTCGAGAGAAACCGTACAAATGCAATGATGATAAAGCTATTTCAGCACAG 580  
 QY 120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGACAGTGA 179  
 Db 581 CATTCTACAATAATACATAAAGAACACATAGTGGAGAGAAACCCCTATGAATGTA-----A 634  
 QY 180 CGATTGTGATAAAGCATTTTTCAGAAATACATTGGAACACATATTGTATCACATTC 239  
 Db 635 TCATGTGTTAAGCCCTTTACACAAACACAGTCATCTCAAAATACATATGTTTACATAC 694



QY 544 ACTTGTCAGTATTACAAATTCATATATAAACAACCTGCATCCAAACTTAATTAATGTCCTTAA 603  
Db 1769 AAATCTTACTGAAACATAAGAAATTCATCTGCGAGAGAAACCCCTA---TGAATGTGAAAAA 1825  
QY 604 TGTGGTAAAGGTTGTGTGGGAAAAAGGTTATCTTTCATCATATGTTAAGTTCATGATGA 662  
Db 1826 TGTGGCAAGCTTTTAACCAAGTCTCTCAATCTTACTAGACATAAGAAAAAGTTCATACAGA 1884

RESULT 5  
AUI23448  
LOCUS  
DEFINITION  
AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA  
sequence.  
AUI23448  
ACCESSION  
VERSION  
AUI23448.1 GI:10948164  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 745)  
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,  
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and  
Isogai, T.  
HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,  
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,  
Suzuki, Y., Sugano, S., Isogai, T.)  
Unpublished (2000).  
CONTACT: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
FEATURES  
source  
1..745  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/c1ones="NT2RM2000319"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/clone\_lib="NT2RM2"  
/notes="vector: pWE18SPL3; mRNA from uninduced NT2 neuronal  
precursor cells"

ORIGIN  
Query Match 7.9%; Score 97.6; DB 9; Length 745;  
Best Local Similarity 50.2%; Pred. No. 2.1e-08;  
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;  
QY 7 GAAAGTACGAAACCAAAATCGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66  
Db 87 GAATGTGGCAAACTTTTAACCGATCTCTCACTCTTACCTACCCATAAGATAATTCATCT 146  
QY 67 AAAAGTATATTTGCATATGAAAGGTTGATAAAGCTTATATCGACCATCATTTATTA 126  
Db 147 GGAGAAACCCCTACAAATGTAAGATGTGGTAAAGCTTTTAACCGTTCTTCAACCTT 206  
QY 127 GAGCAACATTTAAGAACCCACGATTAATGATCGACCGTATAAATGATACAGTGGAGATTGT 186  
Db 207 ACTACCCATAGAAAATTCATCTAGGAGAGAAACCTTACAAATGT-----GAAGATGT 260  
QY 187 GATAAAGCATTTTTCAGAAAATTCATCTAGGAGAAACATATTTGATCATCTCCGAAAAA 245  
Db 261 GGCAAGGCTTTTAAGCAGTCTCTCAACCTTACTACATAAGATAATTCATCTGGAGAG 320

QY 247 AACCACTTCATTTGTTGAGTGTGGTAAAGGGGTTAAATTTCTGACACACACTTTGAAAGA 306  
Db 321 AACCCCTACAAATGTAAATAATGTGAAAGAGCCCTTTAACCCAGTCTGCACACCTTACTACA 380  
QY 307 CATGAATTCACCCATACAAAGTCAATTTAAATGTACATTTGAAATTTCTCAAGAGGATTT 366  
Db 381 CATGAGGTAATTTCTACTTGGAGAGAAACCCCTACAAATGTGAAATAATGTGAAAGAGCCCTTT 440  
QY 367 TATAA---ACATCAATCTTTAAGACATCATATATTTCTGTTTCATGAAAAAACAATTAACG 423  
Db 441 AATCATTTCTCACACCTTACTACATACATAAGATAATTTCTACTGAGAGAAACCTTACAAA 500  
QY 424 TGTAAACAATGATAAAGTTTTTCACTGACCTTCAAAATTTAGCACACATATAAATTAATA 483  
Db 501 TGTAAAGAAATGTGGTAAAGCTTTTAAACACCTCTTCAACCCCTTACTAAACATAAGATAAT 560  
QY 484 CATCATGTTGATCTCTGCTTATCAATGTGATCATCTGTTCTTTTAAAAAATTTTCAAA 543  
Db 561 CATCTGGAGAGAGCCCTTACAAATGTAAAGATGTGAAAGCTTTTAAACCATCTCTCA 620  
QY 544 ACTTGGTCAGTATTACAAATTTTCAATATAAACAACCTGCATCCAAACCTTAATATGTCCTAAA 603  
Db 621 AAACCTTACTGAACATAAGAAATTTTCACTGAGAGAAACCCCTA---TGAATGTGAAAAA 677  
QY 604 TGTGTAAGGTTGTTGGGAAAAAGGTTTATCTTTCATATATGTTAAAGTCAT 657  
Db 678 TGTGCAAGGCTTTTAAACAGTCTCTCAATCTTACTAGACATAAGAAAGTCAT 731

RESULT 6  
BC036394  
LOCUS  
DEFINITION  
BC036394  
ACCESSION  
VERSION  
BC036394.1 GI:23025784  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 2672)  
AUTHORS  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heish, P.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Udén, T.B., Loquellano, N.A., Peters, G.J.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,  
Schnerbach, A., Schain, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL  
MEDLINE  
22388257  
PUBMED  
12477932  
REFERENCE  
2 (bases 1 to 2672)  
AUTHORS  
Straussberg, R.  
Direct Submission  
Submitted (05-AUG-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>



484 CATCATGGTG 493  
543 CATACTGGAG 552

RESULT 8  
B99387  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

B99387  
CIT-HSP-2281A10.TF CIT-HSP 593 bp DNA linear GSS 26-JUN-1998  
genomic survey sequence.

B99387  
B99387.1 GI:3027197  
GSS.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 593)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.B., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
Unpublished (1998)  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Classes are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tcdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1..593  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7147044"  
/db\_xref="taxon:9606"  
/clone="2281A10"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/notes="Vector: pBelobAC11, Site\_1: HindIII; Site\_2:  
HindIII"

ORIGIN  
Query Match 7.4%; Score 91; DB 28; Length 593;  
Best Local Similarity 54.4%; Pred. No. 3.7e-07;  
Matches 231; Conservative 0; Mismatches 185; Indels 9; Gaps 2;  
69 AAGATATATTGGACATATGAAGGTTGTGATGAAGGCTATATCGACCATCATATTAGA 128  
70 |||||  
47 AAGAAACCCCTACAAATGTGAAGATGTGGCAAAAGCTTTTAGCCAGTCTCAACCCCTAG 106  
129 GCACATTTAAGACCCACACAGTATGATCGACCGTATTAATGTACAGTGACGATGTGA 188  
107 AAAACATGAGATATTCATCTACTGGAGAGAAACCCCTACAAATGT-----GAAGATGTGG 160  
189 TAAAGCATTTTTCAGAAAATCACATTTTGGAAAACACATTTGTATCATCATTCGAAAAAAA 248  
161 TAAAGCTTTTAAAGTGGTCTCACACCTTACTAGACATAAAGTAAATTCATCTGAAGAGAA 220  
249 ACCATTCATTTTCAGTGTGTGGTAAAGGGTTAATTCGACACACATTTGAAGAGACA 308  
221 ACCCTACAAATGTGAAGATGTGGCAAGGGTTTAAACCATTTCTAGCCCTTAGGAAACA 280  
309 TGAATCACCACATCAAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTTTA 368  
|||

281 TAAGATAATTCTATCTGGAAGAAACCCCTACAAATGTGAAGATGTGGCAAAAGCTTTAG 340  
369 TAAACATCAATCTTTT---AAGACATCATATATATCTGTTCATGAAAAAACAATTAAAGTG 425  
341 CCAGTCTCTCAACTCTTAGAAAACATGAGATAATTCTACTGAGAGAAACCCCTACAAATG 400  
426 TAAACATGTAATAAAGTTTTCTACTCGACCTTCAAAATAGCAACACATTAATAAACA 485  
401 TGAAGAATGTGGTAAGCTTTTAAGTGTCTCTCAAAATCTTACTGTATACATAGGTAATCA 460  
486 TCATG 490  
461 TACTG 465

RESULT 9  
BX414627  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX414627  
Homo sapiens THYMUS Homo sapiens cdna clone CS0CAP001YD12  
5-PRIME, mRNA sequence.  
BX414627  
BX414627.1 GI:30649903  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1163)  
Li,W.B., Gruber,C., Jesse,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9703.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0CAP001YD12&cluster=9703.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0CAP001YD12062P1.

FEATURES  
source  
Location/Qualifiers  
1..1163  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0CAP001YD12"  
/tissue\_type="THYMUS"  
/clone\_lib="Homo sapiens THYMUS"  
/note="Vector: pCMVSPORT\_6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."

ORIGIN  
Query Match 7.4%; Score 91; DB 13; Length 1163;  
Best Local Similarity 50.1%; Pred. No. 3.5e-07;  
Matches 289; Conservative 8; Mismatches 268; Indels 12; Gaps 3;  
89 AAGGTTGATTAAGCCCTTATATCGACCATCATATTAGACCAACATTTAAGAACCCACA 148  
90 |||||  
121 AAGATGTGGTAAGCTTTTAAACCGATCTTCAACCTTACTACATAGAGRTTCATA 180  
149 GTAATGATCGCCGTATATAATGTACAGTGGACGATGTGTATAAGCATTTTTCAGAAAT 208  
181 CTGGAGAGAAACCTTACAAATGT-----GAAGATGKGGCAAAAGCCCTTAAAGCAGTCT 234  
209 CACATTGGAACACATATTTGATCATCTCCGAAAAAACCATTCCATTGTTCAAGTGT 268  
235 CAACCTTACTACATAGATAATTTCTACTGGAGAAACCCCTACAAATGTAAGAT 294  
|||



```

RESULT 11
BC037782
LOCUS
DEFINITION
Homo sapiens hypothetical protein LOC148206, mRNA (cdna clone
IMAGE:4797729), with apparent retained intron.
ACCESSION
BC037782.1 GI:23349076
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2698)
Strausberg, R.L., Reinsold, E.A., Grouse, L.H., Derge, J.G.,
Klauser, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Spatichenko, L., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahey, J.J., Helton, E., Kettelman, M., Maman, A., Young, A.C., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
23386257
12477932
REFERENCE
2 (bases 1 to 2698)
Strausberg, R.
Direct Submission
Submitted (16-SEP-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smallos, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 70 Row: 9 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein
This clone has the following problem: retained intron.
Location/Qualifiers
1..2698
/organism="Homo sapiens"
FEATURES
source

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4797729"
/tissue_type="Brain, hypothalamus"
/clone_lib="NIH_MGC_96"
/lab_host="DH10B"
/notes="Vector: pBluescript"

```

## ORIGIN

```

Query Match      7.3%; Score 90.4; DB 11; Length 2698;
Best Local Similarity 51.8%; Pred. No. 4.1e-07;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY 7 GAAGTGCAGAAACCAATCGATATCATCTTTAATATCTCTTCTTCTTCATCGTCCC 66
DB 1042 GAATGTGGCAAGCCTTCTACATCTTTCACACCTTACTACATAGGTAACTACT 1101
QY 67 AAAAGTATATTTCACATATGAAGGGTGTGATAAGCCTATAATCGACCACTATT 126
DB 1102 GGAGAGAAGCCCTTCAATGTGAAGATGTGTAAGCTTTTAACCACTTCAGCCCT 1161
QY 127 GAGCAACATTTAAGAACCCAGTAATGATCGACCGTATTAATGATGACAGTGG 186
DB 1162 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCACAAATGT-----GA 1215
QY 187 GATAAAGCATTTTTCAGAAAAATCATTGGAACACATATTTGATCATCATTCGAAA 246
DB 1216 GACAAAGCTTTTAAACCGATTCTCATACCTTACTTAACATAAGATAATTCATT 1275
QY 247 AAACCATTCCTAGTGTGTGGTAAAGGGTAAAGGGTAAATCTTCGACACACAT 306
DB 1276 AAATCTTCAAAATGTGAACAATGTGGCAAGGCTTTAACTGGTCTTCAACCC 1335
QY 307 CATGAATACCCCATACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAA 366
DB 1336 CATAGAAGAAATTCATCTGGAGAGAAACCTTCAAAATGTGAAGATGTGGCA 1395
QY 367 TATAAACAAT---CAATCTTTAAGACATCATATATATCTGTTGATGAAAAA 423
DB 1396 AATGTGTCTTCCACACCTTACTACATAAGATGATTCATCTGAGAGAGAAAC 1455
QY 424 TGTAAACAATGTAATAAGTTTTCACTCGACCTTCAAAATTAGCACACATAAA 483
DB 1456 TGTGAAGAAATGTGGCAAGCCCTTAAACCACTCTCAAAACTTACTATACA 1515
QY 484 CATCATGTGGGATCTCCT 501
DB 1516 CATACTGGAGAGAAACCT 1533

```

## RESULT 12

```

BC028252
LOCUS
DEFINITION
Mus musculus, clone IMAGE:3674739, mRNA.
ACCESSION
BC028252
VERSION
BC028252.1 GI:20380246
KEYWORDS
HTC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2253)
Strausberg, R.
Direct Submission
Submitted (08-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: nisc\_mgc@hgrl.nih.gov  
 Akter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Grante, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., McGaspi, R.,  
 Maduro, O.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 65 Row: a Column: 17  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 13386417  
 This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
 1..2253  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CZECH II"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3674739"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor  
 arose spontaneously from a senescent normal mammary  
 (clonal) outgrowth infected with the virus MNV.".  
 /clone\_lib="NCI CGAP\_Lu29"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 7.3%; Score 90.2; DB 11; Length 2253;  
 Best Local Similarity 51.9%; Pred. No. 4.6e-07;  
 Matches 256; Conservative 0; Mismatches 228; Indels 9; Gaps 2;  
 QY 4 AGTGAAGTGAGCAACCAATCGATATCATCTTTTAATATCTCTCTCTTCTTCATCAGT 63  
 DB 823 AATCAATGTGATAAAGCCTTTTCAACACACACAGTCTCTTCAAACTCATAGAAGACACAT 882  
 QY 64 CCCAAAAGTATATTGCGCATATGAGGGTGTGATTAAGCCTATTAATCGACCATCTTA 123  
 DB 883 ACCGGAAGGAACCCCTTCAATGTATCAATGTGATTAAGCCTTTCCCGTAACATGTT 942  
 QY 124 TTAGAGCAACATTTAAGACCCACAGTAAATGATCGACCGTATAAATGTACAGTGGACGAT 183  
 DB 943 GTGCAACCCATATAGGATACATCTCGAGAAACCCCTTCAATGTA-----ATCAA 996  
 QY 184 TGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCATCTCCGAA 243  
 DB 997 TGTGATAAAGCCTTTTTCAGACACAGTACTCTTCAAACTCATAGAAGAACACATACCGGA 1056  
 QY 244 AAAAACCATTCCATTTTCAGTGTGTGTAAGGGTTAATTCGCAACACATTTGAAA 303  
 DB 1057 GAGAAACCCCTTCAATGTATCAATGTGATTAAGCCTTTCTGTTAAACATAGTCTCCAA 1116  
 QY 304 AGCATGAATACCCCATCAAAAGTCATTTAATGTACATTTGAAAAATTTGCAAGAAGCA 363  
 DB 1117 ACACATAGGAATACATCTCGGGAACCAACCCCTACAAATGTAATCAATGTGATAAGCC 1176  
 QY 364 TTTTATACAC---TCAATCTTTAGACATCATATATTATCTGTTCAATGAAAAACATTA 420  
 DB 1177 TTTTTCACACCTTTCACTCCAAATCATATATAGACACATCTAGAGAGATGCTTAC 1236  
 QY 421 ACCTGTAAACATGTAAATTAAGTTTTCCTCGACCTTCAAAATTTAGCACACATAAATTA 480  
 DB 1237 AATGTAACTAATGTGACAAAGCCTTTTACGACATAGAAATCTTTCAGATTCATAGTAGA 1296  
 QY 481 AATCATCATGGTG 493

Db 1297 AAACATACGGAG 1309

## RESULT 13

BE161630  
 LOCUS  
 DEFINITION  
 MR3-HT0446-260300-201-f06 HT0446 Homo sapiens cDNA, mRNA sequence.  
 BE161630  
 ACCESSION  
 BE161630.1 GI:8624351  
 VERSION  
 EST.  
 BE161630.1  
 KEYWORDS  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 604)  
 AUTHORS  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 TITLE  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL  
 PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE  
 20202663  
 PUBMED  
 10737800  
 COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4&t2=MR3-HT0446-260>)  
 300-201-f06&t3=2000-03-26&t4=1  
 Seq primer: puc 18 forward  
 High quality sequence stop: 583.

## FEATURES

Location/Qualifiers  
 1..604  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="HT0446"  
 /note="Organ: head neck; Vector: puc18; Site: 1; SmaI;  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 7.2%; Score 89.2; DB 10; Length 604;  
 Best Local Similarity 53.7%; Pred. No. 8.1e-07;  
 Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;  
 QY 69 AAAGTATATTGACATATCAAGGCTGTGTAAGCCTATAATCGACCATCATTTATTA 128  
 DB 102 AGAGAAACCTTCAAAATGTGAAGATGTGACAAAGTTTTTAGTCGCTATCACACCTGA 161  
 QY 129 GCAACATTTAAGAACCCACAGTAAATGATCGACCGTATAAATGTACAGTGGACGATTGTA 188  
 DB 162 AAGCATAGGAGATTCATCTACTCGAGAGAAACCGTACAAATGTAAGT-----TTGTGA 215  
 QY 189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCATCATTTCCGAAAAAAA 248  
 DB 216 CAAGGCTTTCAGAGCTGATTTCACACCTGGCACCAACATATTGTAATTCACATCGAGAGAA 275

249 ACATTCATGTTCACTGTTGTTAAAGGGTTTAATTTCTCGACACACTTGAAGAAGACA 308  
Db ACCTTACAAGTGAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTTGTATGCA 335  
Qy TGAATCAACCCATAC---AAAGTCATTTAAATGTACATTTGAAAATTTGTCACGAAGCAAT 365  
Db TAAGTCAATTCATCTCGAGAGAAATGTTTCAAGTGTAAATGATGTGGCAAGTTTTAA 395  
Qy TTATAACATCAATCTTTTAAGACATCATATATATATCTGTTTCATGTAAGAAAACATTAACGTG 425  
Db TCACAAATCAACCTTGCTGATCATCATAGACTTTCATCTGAGAGAGAAACCTTACAAAGTG 455  
Qy TAAACAATGTAATAAAGTTTTCTACCTGCACCTTCAAAATTTAGCACACATAAATTAAGAACA 485  
Db TAATGAATGTGGCAAGGTTTTTAATGAAATCAACCTTGAACATCATCATAGAGTTCA 515  
Qy TCATGCGTGGATCTCCT 501  
Db TATCGGAGAAAACCT 531

RESULT 14  
BC032590  
LOCUS  
DEFINITION Homo sapiens cDNA clone IMAGE:5502691, containing frame-shift errors.  
VERSION BC032590.1 GI:21619671  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

BC032590 3609 bp mRNA linear HTC 19-NOV-2003  
DEFINITION Homo sapiens cDNA clone IMAGE:5502691, containing frame-shift errors.  
VERSION BC032590.1 GI:21619671  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaez T.E., Brownstein M.J., Udén T.B., Teohiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E., Jones S.J. and Marra M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE  
JOURNAL MEDLINE  
PUBMED  
REFERENCE  
AUTHORS Strausberg R.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

REMARK  
COMMENT

Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Akhter N., Ayele K., Beckstrom-Sternberg S.M., Benjamin B., Blackesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S., Dietrich N.L., Grifone S., Guan X., Gupta J., Haghighi P., Hansen N., Ho S.-L., Karlins E., Kwong P., Latic P., Legaspi R., Maduro Q.L., Masiello C., Maskeri B., Mastrian S.D., McCloskey J.C., McDowell J., Pearson R., Stantripop S., Thomas P.J., Touchman J.W., Tsurgenev C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and Green E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 69 Row: 9 Column: 20  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13430873  
This clone has the following problem: frame shifted.

FEATURES  
source  
Location/Qualifiers  
1..3609  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5502691"  
/tissue\_type="Eye, retinoblastoma"  
/clone\_lib="NIH\_MGC\_67"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 7.2%; Score 85.2; DB 11; Length 3609;  
Best Local Similarity 53.7%; Pred. No. 6.7e-07;  
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;  
QY 69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCCTATATCGACCATCATTTTGA 128  
Db 587 AGAGAAACCTTACAAATGTGAAGATGTGACAAAGTTTTTGTGCTTAATCACACTTGA 646  
QY 129 GCAACATTTAAGAACCCACAGTAAATCATCGACCGTATAAATGTACAGTGGACGATTGTA 188  
Db 647 AAGACATAGGAGAAATTCATCTGAGAGAAACCGTACAAATGTAAGT-----TGTGA 700  
QY 189 TAAAGCATTTTTCAGAAATACATTTGGAACACATATTTGATCATCATTCGAAAAAAA 248  
Db 701 CAAAGCTTTCAGACGTGATTTCACCTGGCACACATATTTGTAATTCACACTGGAGAGAA 760  
QY 249 ACATTCATCTGTTGAGTGTGGTAAAGGGGTTAAATTCGACACACACTTGAAGAAGACA 308  
Db 761 ACCTTACAAAGTGAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTTGTATGCA 820  
QY 309 TGAATCAACCCATAC---AAAGTCATTTAAATGTACATTTGAAAAATTTCAAGAAGCAAT 365  
Db 821 TAAGTCAATCATCTGAGAGAAATGTTACAAAGTGAATGATGTGGCAAGGTTTTTAA 880  
QY 366 TTATAAACATCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAACATTAACGTG 425  
Db 881 TCACAAATCAAAACCTTGCATGTTCATAGACTTCATCTGAGAGAGAAACCTTACAAAGTG 940  
QY 426 TAAACAATGTAATAAAGTTTTCTACCTGCACCTTCAAAATTTAGCACACATAAATTAAGAACA 485  
Db 941 TAATGAATGTGGCAAGGTTTTTAATGAAATCAACCTTGAACATCATCATAGAGTTCA 1000  
QY 486 TCATGCGTGGATCTCCT 501  
Db 1001 TATCGGAGAAAACCT 1016

RESULT 15  
AL705393  
LOCUS  
DEFINITION DKF20686M1835 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
AL705393  
ACCESSION  
VERSION AL705393.1 GI:19688748

AL705393 535 bp mRNA linear EST 04-SEP-2003  
DEFINITION DKF20686M1835 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
AL705393  
ACCESSION  
VERSION AL705393.1 GI:19688748

KEYWORDS EST. Homo sapiens (human)  
SOURCE ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 535)  
AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Weil, B. and  
Wiemann, S.  
TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and  
Wiemann, S.)  
JOURNAL Unpublished (2001)  
COMMENT Contact: MIPS  
MIPS  
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
This is the 5' sequence of the clone insert.  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de  
sequenced by MediGenomix (Martinsried/Germany) within the cDNA  
sequencing consortium of the German Genome Project. No sl sequence  
available.  
This clone (DKFZ686M1835) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1..535  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZ686M1835"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="686 (synonym: hlccc3)"  
/note="Vector: pTriplex2; Site\_1: SfiI; Site\_2: SfiIB;  
cDNA-collection"

ORIGIN  
Query Match 7.2%; Score 88.6; DB 9; Length 535;  
Best Local Similarity 53.3%; Pred. No. 1.le-06; Indels 9; Gaps 2;  
Matches 237; Conservative 0; Mismatches 199;  
QY 60 ACCTCCCAAAAGTATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119  
DB 86 ACATCTGGGAAGAATCTTTCAATGTAAAGAAAGTCAATTTGCATGCTTTC 145  
QY 120 ATTATTAGAGCAACATTTAAGAACCCACAGATATGATCGACCGTATAATGTACAGTGA 179  
DB 146 ACACCTTAGCTCAACATAAAGAAATTCATAGTGAGAGAAACCCCTACAATGTAAA----- 200  
QY 180 CGATTGTGATAAGCATTTTTCAGAAATCAATTTGGAAACACATATTTGTATCATTC 239  
DB 201 -GAATGTGGGAAGCCTATAATGAGACCTCAACCTTTCTACATATAAGAAATTCATAC 259  
QY 240 CGAAAAAAACCATTCATTTGCTAGTGTGTGTAAGGGTTAATTCGACACACTT 299  
DB 260 TGGAAAGAAACCCCTCAAAATCGAGAGTGTGAAAGCCTTTAACCCTCTCACACT 319  
QY 300 GAAAGACATGAATCACCCTACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359  
DB 320 TACTACATATGATATTTCTACTGGAAGAAACCCCTACAATGTGAGAGTGTGGCA 379  
QY 360 AGCATTTTATAACAT---CAATCTTTAAGACATCATATATTTATCTGTTCATGAAAAAC 416  
DB 380 AGCTTTTAAACATCTGCAACCTTACTACACATAGAGAATTCATCTGGAGAGAAC 439  
QY 417 ATTAACTGTAAACAATGTAATAAGTTTTCACCTCGACCTTCAAAATTAGCACACATAA 476  
DB 440 CTACAAATGTGAAGAATGTGGCAGAGCTTTTACCCAGTCTCTAACCCCTTACTGCACATAA 499  
QY 477 ATTAAACATCATGGTGGATCTCCT 501  
DB 500 GATAATTCATGCTCGAGAGAACCT 524

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:49:53 ; Search time 5088.16 Seconds  
(without alignments)  
10554.310 Million cell updates/sec

Title: US-09-831-804-2  
Perfect score: 1239  
Sequence: 1 atgagtgaagtgaagaaac.....catcagtgattcttcgataa 1239

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_ov.\*
- 5: gb\_ov.\*
- 6: gb\_ov.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vt.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_mu.\*
- 21: em\_mu.\*
- 22: em\_mu.\*
- 23: em\_mu.\*
- 24: em\_mu.\*
- 25: em\_mu.\*
- 26: em\_mu.\*
- 27: em\_mu.\*
- 28: em\_mu.\*
- 29: em\_mu.\*
- 30: em\_mu.\*
- 31: em\_mu.\*
- 32: em\_mu.\*
- 33: em\_mu.\*
- 34: em\_mu.\*
- 35: em\_mu.\*
- 36: em\_mu.\*
- 37: em\_mu.\*
- 38: em\_mu.\*
- 39: em\_mu.\*
- 40: em\_mu.\*
- 41: em\_mu.\*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1239	100.0	1239	6	BD274350	BD274350 Candida a
2	1239	100.0	2060	6	BD274349	BD274349 Candida a
3	1235.8	99.7	1239	6	AX489193	AX489193 Sequence
4	217.4	17.5	1560	8	VSCNOFEAT	M90638 Saccharomyc
5	217.4	17.5	1739	8	YSCFIIIA	M80611 Saccharomyc
6	217.4	17.5	37497	8	YSCP6677	U95841 Saccharomyc
7	199.2	16.1	2133	8	YSCRF026A	M33924 S.cerevisia
8	193	15.6	797	11	CNS06ING	AL400514 T7 end of
9	101	8.2	169063	9	AC138126	AC138126 Homo sapi
10	101	8.2	177299	9	AC073544	AC073544 Homo sapi
11	99.4	8.0	2132	6	AX714042	AX714042 Sequence
12	99.4	8.0	2132	9	AX056088	AX056088 Homo sapi
13	99.4	8.0	136227	9	AC098500	AC098500 Homo sapi
14	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
15	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
16	97.8	7.9	1203	9	D70831	D70831 Homo sapien
17	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
18	97.2	7.8	115995	9	AC011494	AC011494 Homo sapi
19	96.6	7.8	2320	6	AR270491	AR270491 Sequence
20	96.6	7.8	2320	9	HSU35376	U35376 Human repre
21	96.6	7.8	68304	10	EX000432	EX000432 Mouse DNA
22	95	7.7	981	9	HSRZF9	X78932 H.sapiens H
23	94.6	7.6	186233	9	AC092329	AC092329 Homo sapi
24	94.6	7.6	189317	2	AC024483	AC024483 Homo sapi
25	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
26	93.6	7.6	1389	9	M27879	M27879 Homo sapien
27	93.6	7.6	15439	9	AC008626	AC008626 Homo sapi
28	93.4	7.5	19585	10	AC124426	AC124426 Mus muscu
29	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
30	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
31	92	7.4	2873	9	HUMFLK	M55422 Human Kruep
32	92	7.4	2873	11	G28705	G28705 SWSS3976 Er
33	92	7.4	2873	11	HUMSWS1269	G18281 human chrom
34	92	7.4	156835	9	AC010620	AC010620 Homo sapi
35	92	7.4	191082	2	AC141066	AC141066 Homo sapi
36	92	7.4	203396	9	AC073210	AC073210 Homo sapi
37	91.8	7.4	1870	9	AY044432	AY044432 Homo sapi
38	91.8	7.4	2110	6	AX748253	AX748253 Sequence
39	91.8	7.4	2110	9	AK093669	AK093669 Homo sapi
40	91.8	7.4	2909	9	HSM08021	BX647875 Homo sapi
41	91.8	7.4	2976	9	BC036439	BC036439 Homo sapi
42	91.8	7.4	3078	6	AX747097	AX747097 Sequence
43	91.8	7.4	3078	9	AK091618	AK091618 Homo sapi
44	91.8	7.4	3123	9	BC040594	BC040594 Homo sapi
45	91.8	7.4	174994	9	AC022145	AC022145 Homo sapi

ALIGNMENTS

RESULT 1	BD274350	BD274350	1239 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD274350	Candida albicans tflIIA gene (CatfIIIA) and the coded CATfIIIA				
DEFINITION	BD274350	protein.				
ACCESSION	BD274350.1	GI:33084118				
VERSION	JP 2002531068-A/2.					
KEYWORDS	Candida albicans					
SOURCE	Candida albicans					
ORGANISM	Candida albicans					
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					
AUTHORS	1 (bases 1 to 1239)					
TITLE	Pallier, F.B., Camier, S. and Sentenac, A.					
	Candida albicans tflIIA gene (CatfIIIA) and the coded CATfIIIA					



```

CC      Key      protein      Location/Qualifiers
FH      source      1..2060
FT      /organism="Candida albicans"
FT
FEATURES
    source      1..2060
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /db_xref="taxon:5476"
ORIGIN
    Query Match      100.0%; Score 1239; DB 6; Length 2060;
    Best Local Similarity 100.0%; Pred No. 1.5e-190;
    Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1      ATAGTGAAGAGTGAAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
DB      720      ATGAGTGAAGAGTGAAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779
QY      61      CGTCCCAAAAGATATTTCCACATATGAAGGTGTGATAAGCCCTATATCGACCATCA 120
DB      780      CGTCCCAAAAGATATTTCCACATATGAAGGTGTGATAAGCCCTATATCGACCATCA 839
QY      121      TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
DB      840      TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899
QY      181      GATTGTGATAAAGCATTTTTCAGAAATCATCTTTGAAACACATATTTGATCACAATTC 240
DB      900      GATTGTGATAAAGCATTTTTCAGAAATCATCTTTGAAACACATATTTGATCACAATTC 959
QY      241      GAAAAAAACCATTCCTAGTGTGTGTAAGGGGTAAATCTCGACAAACCTTG 300
DB      960      GAAAAAAACCATTCCTAGTGTGTGTAAGGGGTAAATCTCGACAAACCTTG 1019
QY      301      AAAAGACATGAATCACCACATCAAGTCATTTAATGTACATTTGAAATTTGCAAGA 360
DB      1020      AAAAGACATGAATCACCACATCAAGTCATTTAATGTACATTTGAAATTTGCAAGA 1079
QY      361      GCATTTTATAACATCAATCTTTAAGACATCATATATTTCTGTTTCATGAAAAACATTA 420
DB      1080      GCATTTTATAACATCAATCTTTAAGACATCATATATTTCTGTTTCATGAAAAACATTA 1139
QY      421      ACGTGTAAACATGTATAAGTTTTCACATCGACCTTCAAAATAGCACAACATAATTA 480
DB      1140      ACGTGTAAACATGTATAAGTTTTCACATCGACCTTCAAAATAGCACAACATAATTA 1199
QY      481      AACCATCATGTGTGATCTCTGCTTATCAATGTGATCATCTCTGTTGTTTAAAAATTC 540
DB      1200      AACCATCATGTGTGATCTCTGCTTATCAATGTGATCATCTCTGTTGTTTAAAAATTC 1259
QY      541      CAAACTTGGTCAGTATTACAAATTTTCATATAAAACACTGCATCCAAAACCTTAAATGTCT 600
DB      1260      CAAACTTGGTCAGTATTACAAATTTTCATATAAAACACTGCATCCAAAACCTTAAATGTCT 1319
QY      601      AAATGTGGTAAAGGTGTGTTGGAAAAAGGTTATCTTTCACATATGTTAAGTCATGAT 660
DB      1320      AAATGTGGTAAAGGTGTGTTGGAAAAAGGTTATCTTTCACATATGTTAAGTCATGAT 1379
QY      661      GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGAAATTTGCAAG 720
DB      1380      GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGAAATTTGCAAG 1439
QY      721      AAAAATGAATTAGTGAACATTAATAATCTTCCATGATGGTAAATATCCCTGATGATTA 780
DB      1440      AAAAATGAATTAGTGAACATTAATAATCTTCCATGATGGTAAATATCCCTGATGATTA 1499
QY      781      TTAAGGAACTGAAGTGAAGAAATTTAGAGAACCTATTAGATCAAGGATCGAAATTAAT 840
DB      1500      TTAAGGAACTGAAGTGAAGAAATTTAGAGAACCTATTAGATCAAGGATCGAAATTAAT 1559
QY      841      AATTTCATGAATTTAGAAACAGAAATTTAAAGTGGAAAGAGATGAAGAGATGAAGAA 900

```

```

DB      1560      AATTTGCATGAATTTAGAAACAGAGAAATTTAAAGTGGGAAGATGAAGAGATGAAGAA 1619
QY      901      GATAGTCTAGATGAAGAAAGAGATGATGTAGATCAGACTCAATCTCAGCTCAAGATCA 960
DB      1620      GATAGTCTAGATGAAGAAAGAGATGATGTAGATCAGACTCAATCTCAGCTCAAGATCA 1679
QY      961      ATAAAAATCATTTACTGCTTCTTTGGAAGGTTCAAGAGAGTGTCTTAAACTTATTCTGAAT 1020
DB      1680      ATAAAAATCATTTACTGCTTCTTTGGAAGGTTCAAGAGAGTGTCTTAAACTTATTCTGAAT 1739
QY      1021      AGTGGGAAGAGATCAATTTGCTTAAAGAAATTAATTTGTGATAGAAATGTTTCTAGAGATAT 1080
DB      1740      AGTGGGAAGAGATCAATTTGCTTAAAGAAATTAATTTGTGATAGAAATGTTTCTAGAGATAT 1799
QY      1081      GATTACCTCGACATTTGAAATGCGATGATCAATTAATTTACAAAGAAATTTGAGTCATCTTA 1140
DB      1800      GATTACCTCGACATTTGAAATGCGATGATCAATTAATTTACAAAGAAATTTGAGTCATCTTA 1859
QY      1141      AATAGTATAGAAAAAGAAAGAAATCCAGAGGTGAACCAATTTGTTAAAAAGCCAGGATG 1200
DB      1860      AATAGTATAGAAAAAGAAAGAAATCCAGAGGTGAACCAATTTGTTAAAAAGCCAGGATG 1919
QY      1201      GATTTATTGCCAAATGAACATCAGTGATTTCTCGATAA 1239
DB      1920      GATTTATTGCCAAATGAACATCAGTGATTTCTCGATAA 1958
RESULT 3
LOCUS      AX489193      1239 bp      DNA      linear      PAT 16-AUG-2002
DEFINITION      Sequence 6493 from Patent WO02053728.
ACCESSION      AX489193
VERSION      AX489193.1      GI:22323205
KEYWORDS
SOURCE      Candida albicans
ORGANISM      Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1      Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
    Gene disruption methodologies for drug target discovery
    Patent: WO 02053728-A 6493 11-JUL-2002;
    Elitra Pharmaceuticals, Inc. (US)
FEATURES
    source      1..1239
    /organism="Candida albicans"
    /mol_type="unassigned DNA"
    /db_xref="taxon:5476"
ORIGIN
    Query Match      99.7%; Score 1235.8; DB 6; Length 1239;
    Best Local Similarity 99.8%; Pred. No. 5.4e-190;
    Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1      ATGAGTGAAGTGAAGAAACCAAAATCGATATCATCTTTAAATATCTTCTTCTTCATCA 60
DB      1      ATGAGTGAAGTGAAGAAACCAAAATCGATATCATCTTTAAATATCTTCTTCTTCATCA 60
QY      61      COTCCCAAAAGATATTTGACATATCAAGGGTGTGAAGCCCTATAATCGACCATCA 120
DB      61      COTCCCAAAAGATATTTGACATATCAAGGGTGTGAAGCCCTATAATCGACCATCA 120
QY      121      TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
DB      121      TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180
QY      181      GATTGTGATAAAGCATTTTTCAGAAATCACAATTTGGAAACACATATTTGTATCATATCC 240
DB      181      GATTGTGATAAAGCATTTTTCAGAAATCACAATTTGGAAACACATATTTGTATCATATCC 240
QY      241      GAAAAAAACCATTCCTAGTGTGTGTAAGGGGTAAATCTTCGACAAACACTTG 300
DB      241      GAAAAAAACCATTCCTAGTGTGTGTAAGGGGTAAATCTTCGACAAACACTTG 300

```



```

RESULT 5
YSCFIIIA
LOCUS      YSCFIIIA               1739 bp      DNA      linear      PLN 13-SEP-1996
DEFINITION Saccharomyces cerevisiae transcription factor IIIA (TFIIIA) gene,
complete cds.
ACCESSION M80611
VERSION   M80611.1 GI:172902
KEYWORDS  transcription factor IIIA; zinc-finger protein; zinc-finger
transcription factor.
SOURCE    Saccharomyces cerevisiae (baker's yeast)
ORGANISM  Saccharomyces cerevisiae
          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
          Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 1739)
AUTHORS   Archambault, J., Milne, C.A., Schappert, K.T., Baum, B., Friesen, J.D.
          and Segall, J.E.
TITLE     The deduced sequence of the transcription factor TFIIIA from
          Saccharomyces cerevisiae reveals extensive divergence from Xenopus
          TFIIIA
JOURNAL   J. Biol. Chem. 267 (5), 3282-3288 (1992)
MEDLINE   92147684
PUBMED   1737784
FEATURES
     source
     1..1739
         Location/Qualifiers
             /organism="Saccharomyces cerevisiae"
             /mol_type="genomic DNA"
             /strain="W303-1A"
             /db_xref="taxon:4932"
             /chromosome="XVI"
             234..1523
                 /gene="TFIIIA"
             234..1523
                 /gene="TFIIIA"
             /notes="zinc finger containing protein"
             /codon_start=1
             /product="transcription factor IIIA"
             /protein_id="AAS08014.1"
             /db_xref="GI:172903"
             /translation="MGGEVLNNEGMPLEAKQETIPISSESESLNLTSTRSSSN
             RPTVECDVGDCAKTRPSILTEHQLSVHQIRAFQCDKCAKSFVKSHLERHLYTH
             SDTPKFCNKGKGVTRQOLKHEVHTKSFCEPGCNLFYKHPQLRAHLISVHL
             HKTCPCNKSFPYRLNHLKSHHDPEVNFYQCTFAGCKMPSRKHDLTHYGS
             HPKLCPIKSPKCVGSLQMWI.IHDDSLVTKWVKCHICPDMSPSRKHLTHYGS
             HTBEDPLEKIKISDIQLVQDHGVQLGNSHNEQDEKISNLRKPKLTNNV
             EFLONVLEKRLSENGENLNLNTVGRKICFYNNCSRTFATKEKYEKHKDKHV
             ELKLIKLEKENKTLVDQNHKEPFIQKETSAGDK"
ORIGIN
Query Match      17.5%; Score 217.4; DB 8; Length 1739;
Best Local Similarity 59.4%; Pred. No. 9.9e-26;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

Qy 62 GTCCAAAAAGTATTTCACATATGAGGGTGTGATAAAGCCTATATCGACCATCAT 121
Db 367 GGCCAAAGACATATTTCTGTGACTATGATGCTGTGATAAGGCAATTTACAAGACCTTCAA 426

Qy 122 TATTAGACACATTTTAAGACCACAGTAATGATGACCGTATAATGTACAGTGGAC 181
Db 427 TTTTGACTGACACCAATTAAGCGTACATCAGGTTTAAA---AGAGATTTCACTGTGATA 483

Qy 182 ATTTGTGATAAGCATTTTTCAGAAAAATCACATTTGGAAACACATATTTGATATCATTCG 241
Db 484 AGTGTGCAAAATCTTCGTTTAAAGAGAGTCACTTAGAGACACTTGTATACGCAATTCG 543

Qy 242 AAAAAAACCATTTCCATTTGTCAGTGTGTGTAAGGGTTAATTTCTCGACACACTTGA 301
Db 544 ATACGAAACCATTTCCATTTGTTCTTATTGTGAAAGAGGATGACGACTCGCCAGCACTGA 603

Qy 302 AAAGACATGAATCACCACATACAAAGTCATTAAATGTACATTTGAAAAATTTGCAAGAAG 361
Db 604 AGCGACGAGAGTAAACGATACCAATCTTTCATTTGTCGAGGAGGATGCAACCTCC 663

```

```

Qy 362 CATTTTATAACATCAATCTTTTAAGACATCATATATATCTGTTCATGAAAAACAATAA 421
Db 664 GATTCTACAGCATCCACAATTAAGGSCACATATTTTATCTGTTCAITTTACATAAACTAA 723

Qy 422 CGTGTAACAATGTAATAAAGTTTCTACTCGACTTCAAAATTTAGCACACATAAAATTA 481
Db 724 CCTGTCCACACTGCAATANAAGCTTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 783

Qy 482 AACATCA-----TGGTGGATCTCTCTCTATCAATGTGATCATCTCTGTTGTTTAAAA 535
Db 784 AACATCAGATCCTGAGGTAGAAATCTTTACCAATGTACTTTTGTCTGTTGTTGCAAG 843

Qy 536 ATTTCCAAACTTGTGTAGTATTAACAATTCATATAAAACAACTGCATCCAAACTTAAT 595
Db 844 AGTTTCGATATGTGTCAATTCGAATCGCATATAAAAATATGATCATCCCTAAATTAAT 903

Qy 596 GTCCATAATGTGTAAAGGTTGTCTGGGAAAAGCTTTATCTTCCATATGTTAAGTC 655
Db 904 GTCTATTTGTAGCAACCATGTGTGGGGAATATGTTTACAATGCACATGATTATTC 963

Qy 656 ATGATGATTTACCATGATCAAAATATGAGCTTGTGATTAATTTGTGATGTGGGAAATTTG 715
Db 964 ATGACGACTCACTAGTAACCAAAAATTTGGAAGTGTCTATATGTCCTGATATGTCCTTTT 1023

Qy 716 CAAAGAAAATGAATTAGTTGACATTAT 744
Db 1024 CTAGAAAACATGATCTTCTCAGGCTAT 1052

RESULT 6
YSCP9677/c
LOCUS      YSCP9677               37497 bp      DNA      linear      PLN 01-AUG-1997
DEFINITION Saccharomyces cerevisiae chromosome XVI cosmid 9677.
ACCESSION U25841 U00094
VERSION   U25841.1 GI:786295
KEYWORDS
SOURCE    Saccharomyces cerevisiae (baker's yeast)
ORGANISM  Saccharomyces cerevisiae
          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
          Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 37497)
AUTHORS   Johnson, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Du, Z.,
          Favell, A., Fulton, L., Gattung, S., Greco, T., Kirsten, J., Kucaba, T.,
          Hallsworth, K., Hawkins, J., Hillier, L., Jier, M., Johnson, D.,
          Johnston, L., Langston, Y., Latreille, P., Le, T., Mardis, E.,
          Meneses, S., Miller, N., Nhan, M., Pauley, A., Peluso, D., Rifken, L.,
          Riles, L., Taich, A., Trevaskis, E., Vignati, D., Wilcox, L.,
          Wohlman, P., Vaudin, M., Wilson, R. and Waterston, R.
          The sequence of a portion of the right arm of Saccharomyces
          cerevisiae chromosome XVI
          Unpublished (1995)
          2 (bases 1 to 37497)
          Miller, N.
          The sequence of S. cerevisiae cosmid 9677
          Unpublished (1995)
          3 (bases 1 to 37497)
          Waterston, R.
          Direct Submission
          Submitted (27-APR-1995) Robert Waterston
          Submitted by:
          Genome Sequencing Center
          Department of Genetics, Washington University,
          St. Louis, MO 63110, USA
          e-mail: mj@sequencer.wustl.edu
FEATURES
     source
     1..37497
         Location/Qualifiers
             /organism="Saccharomyces cerevisiae"
             /mol_type="genomic DNA"

```

NEIGHBORING COSMID INFORMATION:  
This sequence includes nucleotides 1-37497 of cosmid 9677. The  
cosmid on the left is 9705.  
Location/Qualifiers  
1..37497  
/organism="Saccharomyces cerevisiae"  
/mol\_type="genomic DNA"

```

gene
CD5
GLCAVHACGRLEASIKVPDKAIQLRPSHTAQVSKAISLDCVDGEYLESDIKQVCV
EAATESFQIGLVEVLMRCSLIDISQGPLLKSAQVIAKIDTIERKIIISLSEKCNQCV
IYUSQVLRFLPWTESKVLDPVESLVSIFENSQPSGSEEDSDVNIKIDTLDDTDD
NVSIAKCLFILTASKYSVQDKETDIAGTRASYWNCIFGIBELTAFITLKEPQVRDAAL
FAFKKSIQLQSNSTWISQIGTATMDINFRVSWCHFKATALEBKPAKTNTWNLGML
KKOTTEFAQVNLKSLAQSDPSWLMGALILEEQDIIGSSKLFAPHSFILNSGHSK
AAQSMYAKVLENIHNGDDEIDETSVKLTITASIALEQFFKKSPSQALQCALITE
ERLHHENANLEANRLIGLEKPKPQDQDERELENFAIKQGFARHILGNSFELSH
NADLSQIISSEDSKSMKTSIKNHICGLSVFLNPDQDTLQFOBELLSISKDSKHL
WDLAKLYDVGBSDTKEALQELTEYATSGADLLVTLIAAMSILDDKREDSLIIIL
BELKALPLSKQIIDKHKDAPYLIEITIRLFRNDTGQVQWRSYFPNNLKYWRLLD
KNQRRJASGQNKVTAEMSCKLYCESKNLKSIOGMFLPWNVTAKALNECF"
complement(9710..11674)
/gene="RPC3"
/gene="RPC3"
complement(9710..11674)
/gene="RPC3"
/note="P9677.11"
/codon_start=1
/product="74 kD subunit of RNA Polymerase III (Swiss Prot
accession number P3349)"
/protein_id="AAB64619.1"
/db_xref="GI:786307"
/translators="MDELLGSEALSAENOTGSTVESEKLVTPEDVMTISSUEORTLNP
DLFLYKELVLAHLGERAAVIGMLVAGLSVRELVKIDGMVDVSVKTTLVSLTOLR
CKVYLQETASIGRKKTIYYINEEGHILIDLSGLIDEIITOMRVNDEEHKQLVASIV
QNVLSGLSDVEYLSVSDSMSTYISLFLVQCEMGYLQISKLUHYTPIEDMWFLL
QVNSIKINPNLSPLDKKRSQAKNAKNTFAKINKPNELSQLTVDPKTSURIKVTP
YESTYINLDFPMKGRSKQILNLAIRTVGSYTAQYKIALRLTEQSPKTRIDTLTQGT
LQOLEBAKFSQDEABLEVEKPTGLTFAIIDLARHLPALDLRGLSLSRKPSDKNSK
GSNAASI.PSKKLTGDEGVYIPALPAAYKSLSQSGDQTEDEEEDLDADTDPHSA
SLINSHUKILASNFPFLNETKGVYVYPYKMLPVLKSSVVEVYIASTLGRSAMPLF
RCIDKNLVSEKILNLAENKEDRDIRSLASLRINSVEIQEVERTADRASRAVFLF
RCKETHSYNFRQILNENMANLILFKKEKILQENSTILLKKAARDVKGRENELLPSEL
NQLKMWNERELNVTFARLSELLSLWEVQNA"
12014..13120
/gene="QCR2"
12014..13120
/gene="QCR2"
/note="P9677.6"
/codon_start=1
/product="Ubiquinol-cytochrome C reductase core protein 2"
/protein_id="AAB64620.1"
/db_xref="GI:786302"
/translators="MLSAARLPQAGSVERLTVSARDAPTKISTLAVKVGGSRYATK
DGAHLNRFNQNTNRSNALKVRESELGGTFKSTUDREVIITLKATFLKDDLPYVV
NALADLYKTAIDKDFHEUTSVLSPARYDYAVAEQCPKVASADQLYATFERKGLGNPL
YDGVERSVLQIDKDFADVYKTEKNSGVENVEADLKRFDVESLITSTLPAGKSLVSK
SEPKFFGEENRVRFTGDSVAYIGIPVKNASLAQVEVLANYTALSELGSLSSAKKL
DKFTDGLGFLTVRDQDSAVSSNIKKIVADLKKGKGLSPAINYTKLKNVAQVESVSS
PIELNFDVAKDQFKLGKFNYYAVGDVSNLPYDEL"
14493..15410
/gene="P9677.5"
/note="P9677.5"
/codon_start=1
/evidence=not experimental
/protein_id="AAB64621.1"
/db_xref="GI:786301"
/translators="MSSNDSNDTKQHTLDDPTGVDVDAIYIPEQETKHHRFKILSDPT
LRDHFIAVGGFCOTGFLWCAYICINVAHVALVVAAPDSQHPGQLIMIAIGRSGVSV
LSDIFMCFVAGSGGALNPAMSLSCILARAVSPTRCVSVAGMAAGGASGAPGVSF
EVLNANSLGCSPTRGLPLEMPTGTAIILCLITVMTAVEKRETNFMAALPTGISLFIAR
VALTAYTGTGVNPARSIAAUAARYFPHVHWIYWTGLLGSILAWSVQLLIQILDYTTT
YVTAASAASTKEKQAKGETSSSSAAVEV"
complement(11542..16012)
/gene="P9677.12"
complement(11542..16012)
/gene="P9677.12"
/note="P9677.12"
/note="similar in C-terminus to Gentamicin
3'-acetyltransferase of P. aeruginosa (Swiss Prot.
accession number P23181)"

```



Db 256 TCACATTTCTAAACATCACGATCCTGAGTAGAGAAATCCCTTACCAATGTAATTTGCTGG 197  
QY 525 TTGTTTAAAAAATTTCAAACTTGGTCAGTATTACAAATTTTATATAAAACAACTGATCC 584  
Db 196 TTGTTGCAAGAGTTTCGGAATGTTGTCACAAATTCGATCCATATATATATATATATAT 137  
QY 585 AARACTTAAATGCTTAAATGTTGTAAGTGTGTTGGGAAAAAGGTTTATCTTCA 644  
Db 136 TAAATTAATGCTTAAATGTTGTAAGTGTGTTGGGAAAAAGGTTTATCTTCA 77  
QY 645 TATGTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 704  
Db 76 CATGATTAATTCATGACGACTCACTAGTACCAAAATTTGGAAGTGTGATATATGCTGGA 17  
QY 705 GGGGAAATTTGCA 718  
Db 16 TATGCTTTTCTA 3

RESULT 8  
CNS06ING 797 bp DNA linear STS 30-NOV-2001  
DEFINITION T7 end of clone AS0A022B07 of library AS0AA from strain CLIB 533  
of Saccharomyces bayanus, sequence tagged site.

ACCESSION AL400514  
KEYWORDS AL400514.1 GI:12156635  
SOURCE Saccharomyces bayanus  
ORGANISM Saccharomyces bayanus  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 797)  
AUTHORS Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Bolin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durand,P., Lepoint,A., Liorente,B.,  
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Petitier,S.,  
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)

REFERENCE 2 (bases 1 to 797)  
AUTHORS Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,  
Aigle,M. and Durand,P.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 5.  
Saccharomyces bayanus var. uvarum  
JOURNAL FEMS Lett. 487 (1), 37-41 (2000)

REFERENCE 3 (bases 1 to 797)  
AUTHORS Genoscope.  
TITLE Direct Submission  
COMMENT Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
segr@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This STS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Kluyveromyces fragilis. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
the other extremity of this insert.  
KEYWORDS for description of this sequence and for the sequence of  
the other extremity of this insert.

FEATURES  
source Location/Qualifiers  
1. .797  
/organism="Saccharomyces bayanus"  
/mol\_type="genomic DNA"  
/strain="CLIB 533"  
/variety="uvarum"

misc\_feature <2..796  
/note="similar to Saccharomyces cerevisiae ORF YPR186c [ PF1 ; PF1IA (transcription initiation factor) ]"  
/evidence=not\_experimental

ORIGIN  
Query Match 15.6%; Score 193; DB 11; Length 797;  
Best Local Similarity 59.3%; Pred No. 1e-21;  
Matches 347; Conservative 1; Mismatches 233; Indels 6; Gaps 1;  
QY 178 GACGATTGTGTAAGCAATTTTCAGAAATACATTTGGAAACATATTGTATCAT 237  
Db 83 GACAATTGTGCAAAATCCCTTTGTAAGAAGATCATCTAGAGAGGCACTGTTTTCGCAT 142  
QY 238 TCCGAAAAAAACCAATTCATTTTCAGTGTGTGTAAGGGGTTAATTCGCAACAC 297  
Db 143 TCTGATACGAGCCATTCATTTCTTATTGTGCAAGGGGTGCAACTCGACAGCA 202  
QY 298 TTGAAAGACATGAATATCCCATGACAAAGTCATTTAAATGTACATTTGAAATTTGTC 357  
Db 203 TTGAAGCGCCACGAAAGTTACACATACCAATCTTTTGTGTCGGAAGATGGATGTGAT 262  
QY 358 GAAGCATTTTATAAATCATCAATCTTTAAGACATCATATATTATCTGTTCTATGAA 417  
Db 263 CTCGATTTTATAAGCAACCCCTCAATTAAGGGCTCATATCTCTGCCACCTAGACAA 322  
QY 418 TTAACGTGTAACAAATGTAATTAAGTTTTCAGTGTGTGTAAGGGTGTGTTT 477  
Db 323 TTAACGTGTAACAAATGTAATTAAGTTTTCAGTGTGTGTAAGGGTGTGTTT 382  
QY 478 TTAACAAATCATGCTGTTTTCAGTGTGTGTAAGGGTGTGTTT 531  
Db 383 TCTAATCATCATGCTGTTTTCAGTGTGTGTAAGGGTGTGTTT 442  
QY 532 AATAATTTCCAACTTGTGTCAGTATTACAAATTTCAATTAATAAACAACCTGCAAACTT 591  
Db 443 ACAGAGTACCGTATATGTCACGTTGCAATCAATCAATCAATCAATCAATCAATCAAT 502  
QY 592 AATGTCCTTAATGTTGTAAGGGTGTGTTGTAAGGGTGTGTTTTCATCATATGTTA 651  
Db 503 AATGTCCTTAATGTTGTAAGGGTGTGTTGTAAGGGTGTGTTTTCATCATATGTTA 562  
QY 652 AATGTCCTTAATGTTGTAAGGGTGTGTTGTAAGGGTGTGTTTTCATCATATGTTA 711  
Db 563 ATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 622  
QY 712 TTTGCAAGAAATGAATTTAGTTGAACATTTAATATATCTTCCAT 756  
Db 623 TTTGCAAGAAATGAATTTAGTTGAACATTTAATATATCTTCCAT 667

RESULT 9  
AC138126/c  
LOCUS AC138126  
DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.  
ACCESSION AC138126  
VERSION AC138126.1 GI:27151357  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 169063)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 169063)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission

JOURNAL Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
FEATURES Location/Qualifiers  
source 1. 169063  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="19"  
/clone="RP11-274A19"  
ORIGIN  
Query Match 8.2%; Score 101; DB 9; Length 169063;  
Best Local Similarity 50.8%; Pred. No. 2e-07;  
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;  
QY 7 GAAAGTGAGCAACCAATCGATATCATCTTTAAATATCTTCTTCTTCATCAGTCCC 66  
Db 80134 GATGTGGCAAGCTTTTAAACGAGCTCTCAATCTTACTACACATAAGAGATTCATCT 80075  
QY 67 AAAAAATATTTGCAATATGAGGGGTGTATAGCCCTATATCGACCATCATTTATTA 126  
Db 80074 GGAGAGAAATCCTACAAATGTGAAGATGTGCAAGCTTTCTATCGATCTCCAAACCT 80015  
QY 127 GAGCAACATTTAAGACCCACAGTATGATCGACCGTATATAATGTACATGACGATGT 186  
Db 80014 ACTGAACATAAGAAATTCATCTGAGAGAAACCTTACATGT-----GAAGATGT 79961  
QY 187 GATAAGCATTTTTCAGAAATACATTTGGAAACACATATTTGATATCATTTCCGAAAA 246  
Db 79960 GGCAGAGCCTTTAAACCATCTCTCACATCTTGCTACACATAAGTAAATTCATCTGGAG 79901  
QY 247 AACCATTCCATGTTTCAGTGTGTGTAAGGGTTAAATTCGCAACACATTTGAAAA 306  
Db 79900 AAACCTTACCATGTGAAGATGTGTAAGCCCTTTAACGAGCTCTCACACCTTACTAGA 79841  
QY 307 CATGAATCAACCCATCAAAAGTCATTTAAATGTACATTTTCAAAATTTGTCAAGAACATTT 366  
Db 79840 CATAAGAGATTCATCTAGGAGAGAAACCTTACCAATGTGAAAAATGTGGCAAGCTTTT 79781  
QY 367 TATAACATCAATCTTTAA---CACATCATATTTATCTGTCATGAAAACATTAACG 423  
Db 79780 AACCAGTCTCAACCTTACTGGACATAGAAAATTCATCTGGTGAAGAACTCTACAAA 79721  
QY 424 TGTAAACAATGTAATAAGTCTTCACTCGACCTTCAAAATTTAGCAACACATAAATTA 483  
Db 79720 CCTAAGAGATGTAACAGTGATTTGAAAACACATTCAAAGTTTCTTAAACATAAAGAAAT 79661  
QY 484 CATCATGGTGATCTCTGCTTATCAATGTGATCATCTCGTGTGTTTAAATTTCCAA 543  
Db 79660 TATGCTGGTGAGAAATCTTAGAAATGTGAAGATGTGAAGAAACCTTTAAAGTTGTCAC 79601  
QY 544 ACTTGGTCAGTATTACAAATTTTATATAAACAACACTGCATCCAAACT 590  
Db 79600 ACTTGAATGTGCATAGATTAATTCATCTACTAATAAACAACCTTCAAGT 79554  
RESULT 10  
AC073544/c 177299 bp DNA linear PRI 19-JUN-2002  
LOCUS AC073544 Homo sapiens chromosome 19 clone RP11-359H18, complete sequence.  
DEFINITION AC073544  
ACCESSION AC073544  
VERSION AC073544.4 GI:21465367  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 177299)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 177299)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 177299)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jun 19, 2002 this sequence version replaced gi:13699752.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.2.  
NOTE: BACTERIAL TRANSPOSONS excised at 14620 and 98099.  
FEATURES Location/Qualifiers  
source 1. 177299  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="19"  
/clone="RP11-359H18"  
ORIGIN  
Query Match 8.2%; Score 101; DB 9; Length 177299;  
Best Local Similarity 50.8%; Pred. No. 2e-07;  
Matches 298; Conservative 0; Mismatches 280; Indels 9; Gaps 2;  
QY 7 GAAAGTGAGCAACCAATCGATATCATCTTTAAACGAGCTCTCAATCTTACTACATAAGAGATTCATCT 163681  
Db 163740 GAATGTGGCAAGCTTTTAAACGAGCTCTCAATCTTACTACATAAGAGATTCATCT 163681  
QY 67 AAAAAATATTTTGCATATGAAAGGGTGTGATAAAGCCTATATCGACCATCATTTATTA 126  
Db 163680 GGAGAGAAATCCTCAAAATGTGAAGATGTGGCAAGCTTTCTATCGATCTCCAAACCT 163621  
QY 127 GAGCAACATTTAAGACCCACAGTATGATCGACCGTATATAATGTACATGAGGATGT 186  
Db 163620 ACTGAACATAAGAAATTCATCTGGAGAGAAACCTTACATGT-----GAAGATGT 163567  
QY 187 GATAAGCATTTTTCAGAAATTCATCTGGAGAGAAACCTTACATGT-----GAAGATGT 163567  
Db 163566 GCAAGAGCTTTTAAACCTCTCTCACATCTTGTCTACATAGTAAATTCATCTGGAG 163507  
QY 247 AAACCATTCATTTGTTCAGTGTGTGTAAGGGTTAAATTCGCAACACATTCGAAAGA 306  
Db 163506 AAACCTTACCAATGTGAAGATGTGTAAGGGTTAAACCTTAAACCTTACTAGA 163447  
QY 307 CATGAATCAACCCATCAAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAGCATTT 366  
Db 163446 CATAAGAGATTCATCTGGAGAGAAACCTTACCAATGTGAAAAATGTGGCAAGCTTT 163387  
QY 367 TATAACATCAATCTTTAA---GACATCATATTTATCTGTCTATGAAAAACATTAACG 423  
Db 163386 AACGAGTCTCAACCTTACTGGACATAGAAAAATTCATCTGGTGAAGAACTCTACAAA 163327  
QY 424 TGTAAACAATGTAATAAGTCTTCACTCGACCTTCAAAATTTAGCAACACATAAATTA 483  
Db 163326 CCTAAGAGATGTAACAGTGATTTGAAAAACATTCAAAGTTTCTTAAACATAAAGAAAT 163267  
QY 484 CATCATGGTGATCTCTGCTTATCAATGTGATCATCTGCTGTTTAAATTTCCAA 543  
Db 163266 TATGCTGGTGAGAAATCTTAGAAATGTGAAGATGTGAAGAAACCTTTAAAGTTGTCAC 163207  
QY 544 ACTTGGTCAGTATTACAAATTTTATATAAACAACCTGCACTCCAAACT 590

2

Estimated Total Number of Errors is 0.1. NOTE: This insert is not the entire sequence of the clone (entire sequence is 153.8kb). It is clipped at the overlap with AC008981. The number of bases overlapped is 27192.	
FEATURES	Location/Qualifiers
source	1. 138627 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="19" /clone="RP11-209J6"
ORIGIN	
Query Match	8.0%; Score 99.4; DB 9; Length 138627;
Best Local Similarity	50.6%; Pred. No. 3.9e-07;
Matches 297; Conservative	0; Mismatches 281; Indels 9; Gaps 2;
QY	7 GAAAGTGCAGAAACCAATCGATATCATCTTTAATATCTTCTCTTCTTCATCAGTCCC 66
Db	1384 GAATGTGGCAAGCTTTTAAACGAGTCTCTCAATCTTCTTACATCAAGAGATTCATCT 1443
QY	67 AAAAGTATATTTGACATATGAAGGGTGTGATAAGCCTATATTCGACCATCATATTATTA 126
Db	1444 GGAGAGAAATCTACAAATGTGAAGATGTGCAAGCTTTCTATCGATCTCTCAAAACTT 1503
QY	127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATATATGTACAGTGGACGATGT 186
Db	1504 ACTGAACATAAGAAATTCATCTGGAGAGAAACCTCTACATGT-----GAAGATGT 1557
QY	187 GATAAAGCATTTTTCAGAAAAATCACATTTTGGAAAACACATATTTGTATCATCTCCGAAAAA 246
Db	1558 GGCMAAGCCTTTAAACCACTCTCCACACCTTGTCTACACATAAGGTAATTCATCTCGAGAG 1617
QY	247 AAACCATTCATTTGTCAGTGTGTGTAAGGGGTAAATCTCGACCAACACTTGAAGA 306
Db	1618 AAACCCCTACCAATGTGAAGAATGTGTAAGGCTTTTAACCACTCTCCACCTTACTAGA 1677
QY	307 CATGAATCACCCATCAAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGACATTT 366
Db	1678 CATGAAGATTCATCTGGAGAGAAACCTACCAATGTGAAAAATGTGGCAAGCTTTT 1737
QY	367 TATAAACATCAATCTTTAA---GACATCATATATTTATCTGTTTCATGAAAAACATTAACG 423
Db	1738 AACCACTCTCAAACTTTACTGGACATAAGAAATTCATCTGGTGAGAACTCTACAAA 1797
QY	424 TGTAACAATGTAATAAGTTTTCACCTCGACCTTGAATTTAGCAACATATAATTAATA 483
Db	1798 CCTAAAGATGTAAACAGTATTTTGAACACACTTCAAGTTTTTCTAAACATATAAGAAAT 1857
QY	484 CATCATGTGGATCTCTGTTATCAATGTGATCATCTGCTGTTTAAAAAATTTCCAA 543
Db	1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGATGTAAACAAACCTTTAAAAAGTTGTAC 1917
QY	544 ACTTGTGAGTATTAATTTCAATTAACAACTGATCTCAAACT 590
Db	1918 ACTTGTATGTGATAAGATAATTCATCTAAAAAARACCTATAAGT 1964
RESULT 13	
AC099500/c	138627 bp DNA linear PRI 29-MAY-2002
LOCUS	AC099500 Homo sapiens chromosome 19 clone RP11-209J6, complete sequence.
DEFINITION	AC099500
ACCESSION	AC099500.2 GI:21240690
VERSION	HTG.
KEYWORDS	Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS	1 (bases 1 to 138627)
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 138627)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3 (bases 1 to 138627)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (29-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On May 29, 2002 this sequence version replaced gi:16930916. Draft sequence produced by DOE Joint Genome Institute www.igi.doe.gov Finishing Completed at Stanford Human Genome Center www.sngc.stanford.edu Quality: Phrap Quality >=40 100% of Sequence;

RESULT 14	
AC138469/c	169500 bp DNA linear HTG 08-JAN-2003
LOCUS	AC138469 Homo sapiens chromosome 19 clone RP11-189C24, *** SEQUENCING IN
DEFINITION	AC138469
ACCESSION	AC138469.1 GI:27544967
VERSION	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
KEYWORDS	Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 169500)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission

247 AAACCATTCATTGTTGAGTGTGGTAAAGGGGTTAAATTTCTCGACACCACTTGAAAGA 306

49433 AAACCCCTACCAATGTGAAGAAATGTGGTAAAGCCCTTTAAACAGTCTCTCACCTTACTAGA 49374

307 CATGAAATCACCCATACAAAGTTCATTTAAATGTACATTTGAAAAATGTCGAAGCAATTTT 366

49373 CATAAGAGAAATTCATCTACTGGAGAGAAACCCCTACCAATGTGAAAAATGTGCAAGCTTTT 49314

367 TATAACATCAATCTTTAA--GACATCATATATATCTGTTCATGAAAAACATTAACG 423

49313 AACGAGTCTCAACCTTACTGGACATAGAAGAAATTCATCTGGTGAAGAACTCTACAA 49254

424 TGTAACAACATGTAATAAGATTTTTCCTCGACCTTCAAAATTAGCACAACTAATTTAAA 483

49253 CCTAAAAGATGTAACAGCTGATTTTGTGAAACACATCTCAAAGTTTCTTAAACATATAAAGAAAT 49194

484 CATCATGTGTGATCTCTGCTTATCAATGTGATCATCTCGTGTGTTTAAAAATTTCCAA 543

49193 TATGCTGTGTGAGAAATCTTAGAAATGTGAAGATGTGAAGAACTTTAAAAAGTTGTAC 49134

544 ACTTGCTCAGTATTACAAATTTTCATATAAAACAATGCATCCAAACT 590

49133 ACTTGATTGTGCTAAGATAATTCATATAAAAAAACCTTATAGT 49087

RESULT 15

LOCUS AK122869

DEFINITION Homo sapiens cDNA FUJ16502 f1s, clone FEBRA2006664, moderately similar to Zinc finger protein 43.

ACCESSION AK122869

VERSION AK122869.1 GI:34529067

KEYWORDS oligo capping; f1s (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M., Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Rio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,K., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2237)

AUTHORS Isogai,T. and Yamamoto,J.

TITLE Direct Submission

JOURNAL Submitted (15-JUL-2003) Takao Isogai, FUJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

COMMENT (E-mail:genomics@hri.co.jp, tel:81-438-52-3975, fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (HRI) ; CDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

source

1..2237

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="FEBRA2006664"

/tissue\_type="brain"

/clone\_lib="FEBRA2"

/dev\_stage="fetal"

/note="cloning vector: pME18SFL3"

ORIGIN

Query Match		7.9%;	Score 98.2;	DB 9;	Length 2237;
Best Local Similarity		52.0%;	Pred. No. 1.6e-06;		
Matches 273;	Conservative	0;	Mismatches 243;	Indels	9; Gaps 2;
QY	69	AAAGTATATTGACATATGAAGGGTGTGATAAAGCCTTAATCGACATCATTTATTAGA	128		
Db	1652	AGAGAAATCCTACAAATGTGAAGATGTGGCAAGCTTTCTATCGATCCTCAAACTTAC	1711		
QY	129	GCAACATTTAAGAACCCACAGTAATGATCGCCGTATTAATGTACAGTGGACGATTGTGA	188		
Db	1712	TGAACATAGAAAATTCATCTGGAGAGAAACCTACACATGT-----GAAGATGTGG	1765		
QY	189	TAAAGCATTTTTCAGAAAATCAATTTGGAACACATATTGTATCACATTCGAAAAAAA	248		
Db	1766	CHAAAGCCTTTAACCATTCCTCACCTTGTCTACACATAAGGTAATTCATCTGGAGAGAA	1825		
QY	249	ACCATTTCCATTGTTCACTGTGTGTTAAAGGGTTAATTCCTGCACACACTTGAAAGACA	308		
Db	1826	ACCTTACCAATGTGAAGATGTGGTAAAGCCTTTAACCCAGTCCCTCACCTTACTAGACA	1885		
QY	309	TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTTGTCAAGAAGCATTTTA	368		
Db	1886	TAAGAGAAATTCATCTGGAGAGAAACCTTACCAATGTGAAAATGTGGCAAGCTTTTAA	1945		
QY	369	TAAACATCAATCTTTAA---GACATCATATATTATCTGTTTCATGAAAAAACAATTACGTG	425		
Db	1946	CCAGTCCCTCAAAACCTTACTGGACATAAGAAAATTCATCTGGTGAGAAACTCTACAAACC	2005		
QY	426	TAAACAATGTAATAAAGTTTTCCTCGACCTTCAAAATTAGCACACATAAAATTTAAAAACA	485		
Db	2006	TAAAGATGTAACAGTGAATTTGAAAAACACTTCAAGTTTTCTTAACATATAAGRAATTA	2065		
QY	486	TCATGGTGGATCTCCTGCTTATCAATGTGATCATCTGCTGGTGTGTTTAAAAAATTCCAAAC	545		
Db	2066	TGCTGGTGAGAAATCTTAGAAATGTGAAGAAATGTAAACAAAACCTTTAAAAAGTTGTACAC	2125		
QY	546	TTGGTCAGTATTACAAATTTTCATATAAAACAACTGCATCCAAACT	590		
Db	2126	TTGATTGTGCATAAGATAATTTCATATAAAAAAACCTATAAGT	2170		

Search completed: May 9, 2004, 08:56:51  
Job time : 5094.16 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:47:06 ; Search time 545.661 Seconds  
(without alignments)  
9646.138 Million cell updates/sec

Title: US-09-831-804-2

Perfect score: 1239  
Sequence: 1 atgagtgaagcagcaaac.....catcagtgattcttcgataa 1239

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- N Geneseq\_29Jan04:\*
  - 1: Geneseqn1990s:\*
  - 2: Geneseqn1990s:\*
  - 3: Geneseqn2000s:\*
  - 4: Geneseqn2001as:\*
  - 5: Geneseqn2001bs:\*
  - 6: Geneseqn2002s:\*
  - 7: Geneseqn2003as:\*
  - 8: Geneseqn2003bs:\*
  - 9: Geneseqn2003cs:\*
  - 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1239	100.0	2060	3	AA15398 DNA encod
2	1235.8	99.7	1239	6	ABZ32206 Candida a
3	99.4	8.0	2132	7	ADA53158 Human cod
4	96.8	7.8	1705	8	ACA98970 cDNA enco
5	96.6	7.8	2320	7	ACA56456 Human sig
6	95.2	7.7	4563	5	RA591317 DNA encod
7	94.6	7.6	2597	7	AA151569 Human nuc
8	92.4	7.5	3639	5	AA564586 DNA encod
9	91.8	7.4	2110	9	ADB63624 Human cDN
10	91.8	7.4	2760	9	ADC56695 Human mac
11	91.8	7.4	3078	9	ADB62468 Human cDN
12	90.4	7.3	2026	7	AD55863 Human nuc
13	90.4	7.3	2114	7	ADA53124 Human cod
14	90.2	7.3	2230	9	ADA52931 Human cod
15	89.2	7.2	2064	9	ADC30762 Human nov
16	89.2	7.2	2622	5	AA568872 DNA encod
17	89.2	7.2	2729	4	AAH16178 Human cDN
18	89.2	7.2	3839	6	ABK83826 Human cDN
19	89	7.2	2298	4	AAH16608 Human cDN
20	89	7.2	2305	5	AA566143 DNA encod
21	89	7.2	3020	5	AA592560 DNA encod
22	89	7.2	3502	4	AA157845 Human pol
23	88.6	7.2	2597	7	AA558555 Human nuc

24	88.6	7.2	4227	8	ACA98938 cDNA enco
25	88.4	7.1	1952	9	ADC58104 Zinc fing
26	88.4	7.1	2509	6	AAD31103 Human tra
27	87.6	7.1	1757	6	ABQ81135 TRAF6-inh
28	87.2	7.0	831	6	ABQ55095 Human ova
29	87	7.0	976	5	AA569188 DNA encod
30	86.8	7.0	2662	6	ABQ93353 Human cDN
31	86.8	7.0	6219	7	ACC46324 Human dit
32	86.8	7.0	6316	7	ACC46373 Human dit
33	86.2	7.0	1377	7	ABX34443 Human mdd
34	86.2	7.0	2558	4	AA526690 Human gen
35	86.2	7.0	2558	4	AA526691 Human gen
36	86.2	7.0	2558	7	ABX74039 Human nov
37	86.2	7.0	2558	7	ABX74040 Human nov
38	86.2	7.0	2607	5	AA587125 DNA encod
39	86.2	7.0	2681	7	ABX34772 Human mdd
40	86	6.9	1890	9	ADC30336 Human nov
41	86	6.9	3309	7	ACC46347 Human dit
42	86	6.9	3400	9	ADB62883 Human cDN
43	85.6	6.9	1549	5	AA567562 DNA encod
44	85.6	6.9	2476	7	ACC46407 Human dit
45	84.8	6.8	2239	7	ADA53516 Human cod

ALIGNMENTS

RESULT 1  
AA15398  
ID AA15398 standard; DNA; 2060 BP.  
XX  
AC AA15398;  
XX  
DT 04-SEP-2000 (first entry)  
XX  
DE DNA encoding a transcription factor designated CATFIIIA.  
XX  
KW Transcription factor; CATFIIIA; DNA-binding protein;  
KW ribosomal RNA 5S gene; fungal infection; ss.  
XX  
OS Candida albicans.  
XX  
FH Key Location/Qualifiers  
FT CDS 720..1958  
FT /\*tag= a  
FT /transl\_except= (pos: 1296..1298, aa: Ser)  
FT /transl\_except= (pos: 1734..1736, aa: Ser)  
XX  
PN WO200028037-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 09-NOV-1999; 99WO-PR002739.  
XX  
PR 10-NOV-1999; 99FR-00014147.  
XX  
PA (HMRI ) HOECHST MARION ROUSSEL.  
XX  
XX Bordon-Pallier F, Camier S, Sentenac A;  
XX  
XX WPI; 2000-376549/32.  
XX  
XX P-PSDB; AA593316.  
XX  
XX New nucleic acid encoding Candida albicans transcription factor, useful  
XX e.g. in screening for antimycotic agents and for immunization.  
XX  
XX Claim 4; Page 32-33; 45pp; French.

The present sequence encodes a Candida albicans transcription factor, designated CATFIIIA. The polypeptide is a DNA-binding protein, which is involved in initiating transcription of the ribosomal RNA 5S gene. The polynucleotide is used to screen for its specific inhibitors, potentially useful as antimycotic agents, to raise an antibody response that is

CC protective against fungal infection and to raise antibodies. Such  
CC antibodies, as well as the polypeptides and polynucleotides are used in  
CC compositions for diagnosing and treating fungal infections, e.g. by  
CC detecting polymorphisms and mutations  
XX  
SQ Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 1239; DB 3; Length 2060;  
Best Local Similarity 100.0%; Pred No. 2.5e-234;  
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ATGAGTGAAGTGAAGCAAGCAATCGATATCATCTTTAATATCTCTCTCTTCATCA 60  
DB 720 ATGAGTGAAGTGAAGCAAGCAATCGATATCATCTTTAATATCTCTCTCTTCATCA 779  
  
QY 61 CGTCCCAAAAGTATATTTCACATATGAAGGCTGTGATAAGCCCTATAATCGACATCA 120  
DB 780 CGTCCCAAAAGTATATTTCACATATGAAGGCTGTGATAAGCCCTATAATCGACATCA 839  
  
QY 121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAATGATGAGTGCAC 180  
DB 840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAATGATGAGTGCAC 899  
  
QY 181 GATTGTGATTAAGCAATTTTCAGAAATCACATTTTGGAAACACATATTGTATCATATCC 240  
DB 900 GATTGTGATTAAGCAATTTTCAGAAATCACATTTTGGAAACACATATTGTATCATATCC 959  
  
QY 241 GAAAAAACAATCCATTCCTGCTGCTGTGTAAGGGGTTAATTCGACACACTTG 300  
DB 960 GAAAAAACAATCCATTCCTGCTGCTGTGTAAGGGGTTAATTCGACACACTTG 1019  
  
QY 301 AAAAGACATGAATACCCATACAAAGTCAATTTAAATGTACATTTGAAAAATGTCAAGAA 360  
DB 1020 AAAAGACATGAATACCCATACAAAGTCAATTTAAATGTACATTTGAAAAATGTCAAGAA 1079  
  
QY 361 GCATTTTAAACATCAATCTTTAAGACATCATATATATCTGTCATCGAAAAACATTA 420  
DB 1080 GCATTTTAAACATCAATCTTTAAGACATCATATATATCTGTCATCGAAAAACATTA 1139  
  
QY 421 ACSTGTAAACAATGTATTAAGTGTTCCTACGACCTTCAAAATAGCACACATAAATTA 480  
DB 1140 ACSTGTAAACAATGTATTAAGTGTTCCTACGACCTTCAAAATAGCACACATAAATTA 1199  
  
QY 481 AAACATCATGCTGATCTCTGCTTATCATATGATGATCATCTGCTGTTGTTTAAAAATTC 540  
DB 1200 AAACATCATGCTGATCTCTGCTTATCATATGATGATCATCTGCTGTTGTTTAAAAATTC 1259  
  
QY 541 CAAACTTGGTCAGTATTACAAATTTTCATATAAAACAACTGCATCCAAAACTTTAAATGTCCT 600  
DB 1260 CAAACTTGGTCAGTATTACAAATTTTCATATAAAACAACTGCATCCAAAACTTTAAATGTCCT 1319  
  
QY 601 AATGTGTAAGAGTGTGTGGAAGAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660  
DB 1320 AATGTGTAAGAGTGTGTGGAAGAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379  
  
QY 661 GATTCTACCATGATCAAAATATGCACTTGTGATTTATGATGTTGGGAAATTTGCAAG 720  
DB 1380 GATTCTACCATGATCAAAATATGCACTTGTGATTTATGATGTTGGGAAATTTGCAAG 1439  
  
QY 721 AAAAAATCAATAGTGAACATTAATATCTTCCATGATGCTGAATATCCCTGATGATTA 780  
DB 1440 AAAAAATCAATAGTGAACATTAATATCTTCCATGATGCTGAATATCCCTGATGATTA 1499  
  
QY 781 TTAAGGAAACTGAAGTGAAGAAATTTAGAGAACCTTATTAGATCAAGGATCGAAATTTAAT 840  
DB 1500 TTAAGGAAACTGAAGTGAAGAAATTTAGAGAACCTTATTAGATCAAGGATCGAAATTTAAT 1559  
  
QY 841 AATTGCAATTAATAGAACAGGAAATTTAAAGTGAAGAGATGAGAGATGAGAGAA 900  
DB 1560 AATTGCAATTAATAGAACAGGAAATTTAAAGTGAAGAGATGAGAGATGAGAGAA 1619  
  
QY 901 GATAGTCTAGATGAAAAAGAGTGTGTTAGATCAGACTCAATGTCAGCTCAAGATCA 960

DB 1620 GATAGTCTAGATGAAAAAGAGTGTGTTAGATCAGACTCAATGTCAGCTCAAGATCA 1679  
QY 961 ATAAATCATTTACTGCTCTCTTTTGAAGGTTCAAGAGTGTCTTAACTATTCTGAAT 1020  
DB 1680 ATAAATCATTTACTGCTCTCTTTTGAAGGTTCAAGAGTGTCTTAACTATTCTGAAT 1739  
QY 1021 AGTGGGAAGAGATCAATTTGCTCTTAAAGAAATTAATTTGTATAGAAATGTTTCTAGAGAATAT 1080  
DB 1740 AGTGGGAAGAGATCAATTTGCTCTTAAAGAAATTAATTTGTATAGAAATGTTTCTAGAGAATAT 1799  
QY 1081 GATTTACGTCGACATTTGAAATGGCATGATGATATTTACAAAGAAATGATCTATTCTTA 1140  
DB 1800 GATTTACGTCGACATTTGAAATGGCATGATGATATTTACAAAGAAATGATCTATTCTTA 1859  
QY 1141 AATAGTATAGAAAAAGAAAGAACTCCAGAGGTTGAACCATTTGGTTAAAAAGCCAGGATG 1200  
DB 1860 AATAGTATAGAAAAAGAAAGAACTCCAGAGGTTGAACCATTTGGTTAAAAAGCCAGGATG 1919  
QY 1201 GATTTATGCGCAATTCAGACATCAGTGTCTTCGATAA 1239  
DB 1920 GATTTATGCGCAATTCAGACATCAGTGTCTTCGATAA 1958

RESULT 2  
AB232206

ID AB232206 standard; DNA; 1239 BP.

XX AB232206;

XX 30-JAN-2003 (first entry)

XX Candida albicans essential gene SEQ ID NO 6493.

XX Funus; Yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
XX signal transduction; DNA replication; cell division; growth;  
XX proliferation; Candida albicans; fungicide; antifungal; gene; ss.

XX Candida albicans.

XX MO200253728-A2.

XX 11-JUL-2002.

XX 26-DEC-2001; 2001WO-US049486.

XX 29-DEC-2000; 2000US-0259128P.

XX 20-FEB-2001; 2001US-00792024.

XX 22-AUG-2001; 2001US-0314050P.

XX (ELIT-) ELITRA PHARM INC.

XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;

XX WPI; 2002-566694/50.

XX P-PSDB; ABP73656.

XX Constructing strains for identifying gene products as effective targets  
XX for therapeutic intervention, by inactivating in the strain one allele of  
XX a gene and placing other allele of the gene under conditional expression.

XX Claim 37; SEQ ID NO 6493; 167pp + Sequence Listing; English.

XX The invention relates to constructing (M1) a strain of diploid fungal  
XX cells in which both alleles of a gene are modified, comprising modifying  
XX one allele by insertion or replacement by a cassette having an  
XX expressible selectable marker and modifying other allele by  
XX recombination, of a promoter replacement fragment with a heterologous  
XX promoter, so that expression of the second allele is regulated by the  
XX promoter. (M1) is useful for constructing a strain of diploid fungal  
XX cells in which both alleles of a gene are modified. The diploid fungal  
XX cells having both alleles modified are useful for identifying a gene that  
XX is essential to the survival or growth of a fungus, a gene that  
XX contributes to the virulence and/or pathogenicity of a fungus, a gene

CC that contributes to the resistance of a diploid fungus to an antifungal  
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
 CC and for identifying a therapeutic agent for treatment of a mammalian  
 CC disease. (M1) is useful for identifying a compound which modulates the  
 CC activity of a gene product, preferably enzymatic activity, carbon  
 CC compound catabolism, biosynthetic, transporter, transcriptional,  
 CC translational, signal transduction, DNA replication and cell division  
 CC activity. The method is useful for identifying a compound having the  
 CC ability to inhibit growth or proliferation of *C. albicans* cells and for  
 CC treating infection by *C. albicans*. The present sequence is that of an  
 CC essential *Candida albicans* gene used in the method of the invention.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office

Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 99.7%; Score 1235.8; DB 6; Length 1239;

Best Local Similarity 99.8%; Pred. No. 9.8e-234;

Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAGTGAAGTACGAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCA 60  
 DB 1 ATGAGTGAAGTACGAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCA 60  
 QY 61 CGTCCCAAAAGTATATTGACATATGAAGGCTGTGATAAGCCCTATATCGACCATCA 120  
 DB 61 CGTCCCAAAAGTATATTGACATATGAAGGCTGTGATAAGCCCTATATCGACCATCA 120  
 QY 121 TTATTAGAGCAATTTAAGAACCCAGATGATCGCCGATATAAATGTCAGTGGAC 180  
 DB 121 TTATTAGAGCAATTTAAGAACCCAGATGATCGCCGATATAAATGTCAGTGGAG 180  
 QY 181 GATTGTGATAAAGCATTTTTCAGAAATCATTTGGAACACATATTTGATCATATCC 240  
 DB 181 GATTGTGATAAAGCATTTTTCAGAAATCATTTGGAACACATATTTGATCATATCC 240  
 QY 241 GAAAAAAACCATTCCTAGTGTGTGTTAAGGGGTAAATCTCGACAAACATTC 300  
 DB 241 GAAAAAAACCATTCCTAGTGTGTGTTAAGGGGTAAATCTCGACAAACATTC 300  
 QY 301 AAAAGACATGAATCACCACATCAAGTCAATTTAAATGTACATTTGAAATTTGTAAGAA 360  
 DB 301 AAAAGACATGAATCACCACATCAAGTCAATTTAAATGTACATTTGAAATTTGTAAGAA 360  
 QY 361 GCATTTTATAACATCAATCTTTAAGACATCATATATATCTTCTTCAATGAAAAACATTA 420  
 DB 361 GCATTTTATAACATCAATCTTTAAGACATCATATATATCTTCTTCAATGAAAAACATTA 420  
 QY 421 ACGGTGAACGATGTAATAAGTTTTCACCTCGACCTTCAAAATTAGCAACAATTAATTA 480  
 DB 421 ACGGTGAACGATGTAATAAGTTTTCACCTCGACCTTCAAAATTAGCAACAATTAATTA 480  
 QY 481 AAACATCATGGTGATCTCTGCTTATCAATGTGATCATCTGTTGTTTTAAAAATTC 540  
 DB 481 AAACATCATGGTGATCTCTGCTTATCAATGTGATCATCTGTTGTTTTAAAAATTC 540  
 QY 541 CAAACTTGGTCAGTATTAATTTTCAATATAAACAACCTGATCCAAATCTTAAATGTCT 600  
 DB 541 CAAACTTGGTCAGTATTAATTTTCAATATAAACAACCTGATCCAAATCTTAAATGTCT 600  
 QY 601 AAATGTGGTAAAGTGTGTTGGGAAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660  
 DB 601 AAATGTGGTAAAGTGTGTTGGGAAAAAGTTTATCTTCACATATGTTAAGTCATGAT 660  
 QY 661 GATTCTACCATGATCAAAATATGACCTGTGATTATTTGATGTGGGAAATTTGCAAG 720  
 DB 661 GATTCTACCATGATCAAAATATGACCTGTGATTATTTGATGTGGGAAATTTGCAAG 720  
 QY 721 AAAAATGAATTTAGTTGAACATTAATATATCTTCATGATGGTAAATATCCCTGATGATTA 780  
 DB 721 AAAAATGAATTTAGTTGAACATTAATATATCTTCATGATGGTAAATATCCCTGATGATTA 780

QY 781 TTTAAGGAAACTGAAGTGAAAAAATTAGAGAACTTATAGATCAAGGATCGAAATTTAAAT 840  
 DB 781 TTTAAGGAAACTGAAGTGAAAAAATTAGAGAACTTATAGATCAAGGATCGAAATTTAAAT 840  
 QY 841 AATTGTCATGAATTAGAAACAGAGAAATTTAAAGTGGAGAGAGATGAAGAATGAAGAA 900  
 DB 841 AATTGTCATGAATTAGAAACAGAGAAATTTAAAGTGGAGAGAGATGAAGAATGAAGAA 900  
 QY 901 GATAGTCTAGATGAAAAAAGAAAGTGTGTTAGATCAGATCTCAATGTGTCAGCTCAAGATCA 960  
 DB 901 GATAGTCTAGATGAAAAAAGAAAGTGTGTTAGATCAGATCTCAATGTGTCAGCTCAAGATCA 960  
 QY 961 ATAAATCATTTACTGCTCTTTTGGAGGTTTCAAGAGTGTCTTAAACTTATCTGAAT 1020  
 DB 961 ATAAATCATTTACTGCTCTTTTGGAGGTTTCAAGAGTGTCTTAAACTTATCTGAAT 1020  
 QY 1021 AGTGGGAAGAGATCAATTTCTCTAAAGATAATTTGTGTAGATGATTTTCTAGAGATAT 1080  
 DB 1021 AGTGGGAAGAGATCAATTTCTCTCTAAAGATAATTTGTGTAGATGATTTTCTAGAGATAT 1080  
 QY 1081 GATTAGCTGACATTTGAAATGGCATGATGATAATTTACAAAGATTTGAGTCATTTCTTA 1140  
 DB 1081 GATTAGCTGACATTTGAAATGGCATGATGATAATTTACAAAGATTTGAGTCATTTCTTA 1140  
 QY 1141 AATAGTATAGAAAAAGAGAAACTCCAGAGAGTGAACCATTTGGTTAAAAAAGCCAGATG 1200  
 DB 1141 AATAGTATAGAAAAAGAGAAACTCCAGAGAGTGAACCATTTGGTTAAAAAAGCCAGATG 1200  
 QY 1201 GATTATTCGAATGAACATCAGTCATTTCTCGATAA 1239  
 DB 1201 GATTATTCGAATGAACATCAGTCATTTCTCGATAA 1239

# RESULT 3

ADAS3158

ID ADA53158 standard; cDNA; 2132 BP.

XX ADA53158;

XX 20-NOV-2003 (first entry)

XX Human coding sequence, SEQ ID 726.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

XX Gene therapy; human; secretory protein; membrane proteins; cancer;

XX inflammatory disease; osteoporosis; neurological disease; gene; ss.

XX Homo sapiens.

XX EPI293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI: 2003-395539/38.

XX P-PSDB: ADA54797.

XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 1; SEQ ID NO 726; 205pp; English.

XX CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;  
 Query Match 8.0%; Score 99.4; DB 7; Length 2132;  
 Best Local Similarity 50.6%; Pred. No. 3.3e-10;  
 Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;  
 QY 7 GAAAGTGAACACCAACCAACGATATCATCTTAAATATCTTCTTCTTCATCAGGCCCC 66  
 Db 1384 GATGTGGCAAGCTTTTAAACGATCTCTCAATCTTACTACACATAAGAAATTCATACT 1443  
 QY 67 AAAAAGTATATTTGCATATGAAGGGTGTGATAAGCCTATAATCGACCATCATTTATTA 126  
 Db 1444 GGAGAGAAATCTACAAATGTGAAGAAATGTGGCAAGCTTTCTATCGATCTCAAAACTT 1503  
 QY 127 GACCAACATTTAGAACCCACACAGTATGATCGACCGTATAAATGTACAGTGGAGATGT 186  
 Db 1504 ACTGAACATGAAGAAATTTATCTCGAGAGAAACCCCTACACATGT-----GAAGAATGT 1557  
 QY 187 GATAAAGCATTTTTCAGAAATCACATTTGGAAACACATATTTGTATCATCTCCGAAAAA 246  
 Db 1558 GGCAGAGCCTTTAAACCACTCTCACACCTTGCTACACATAAGGTAATTCATCTGGAGAG 1617  
 QY 247 AAACCATTCATTTGTCAGTGTGTGTAAAGGGTAAATTCGACACACTTGAAGA 306  
 Db 1618 AAACCTTACCAATGTGAAGAAATGTGAAGCCTTTAAACGATCTCACCTTACTAGA 1677  
 QY 307 CATGAATACCCATCAACAGTCATTTAAATGTACATTTGAAATTTGCAAGAAGCATTT 366  
 Db 1678 CATAAGAGAAATTCATCTCGAGAGAAACCCCTACCAATGTGAAATATGTGGCAAGCTTT 1737  
 QY 367 TATAACATCAATCTTAA---GACATCATATATATCTGTCATGAAGAAACATTAACG 423  
 Db 1738 AACAGTCTCAACCTTACTGGACATGAAGAAATTCATCTGGTGAAGAACTCTACAA 1797  
 QY 424 TGTAAACATGTATAAAGTTTCACTCGACCTTCAAAATTTAGCACACATATAAATAA 483  
 Db 1798 CCTAAAGAGTGAACAGTATTTTGAACACACTTCAAGATTTCTAAACATATAAAGAAAT 1857  
 QY 484 CATCATGTGTGATCTCTCTTATCATGTGATCATCTCGTGTGTTGTTTAAATTTTCAA 543  
 Db 1958 TATGCTGTGAGAAATCTTAGAAATGTGAAGAAATGTACAAACCTTTAAAGTTGTAC 1917  
 QY 544 ACTGTGTGATTTACAAATTTTCATATAAACAACCTGCATCCAAACT 590  
 Db 1918 ACTTGATGTGCAAGATAAATTCATATAAATAAATAAATAAATAAATAAATAAATAAATAA 1964  
 RESULT 4  
 ID ACA98970 standard; cDNA; 1705 BP.  
 XX AC ACA98970;  
 XX AC ACA98970;  
 DT 25-JUL-2003 (first entry)  
 XX DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.  
 XX KW Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;  
 KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiallergic; antiinflammatory; thymimetic; gene therapy;  
 KW cell proliferative disorder; cancer; atherosclerosis;  
 KW neurological disorder; epilepsy; Huntington's disease; stroke;  
 KW immune disorder; inflammatory disorder; AIDS; allergy;  
 KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;  
 KW protein-protein interaction; drug-target interaction;  
 KW gene expression profile; gene; ss.

XX OS Homo sapiens.  
 XX WO2003023003-A2.  
 XX 20-MAR-2003.  
 XX 05-SEP-2002; 2002WO-US028540.  
 XX 07-SEP-2001; 2001US-0317792P.  
 XX 14-SEP-2001; 2001US-0322270P.  
 XX 21-SEP-2001; 2001US-0324040P.  
 XX 28-SEP-2001; 2001US-0326732P.  
 XX 19-OCT-2001; 2001US-0346718P.  
 XX 25-JAN-2002; 2002US-0351749P.  
 XX 22-FEB-2002; 2002US-0359498P.  
 XX (INCY-) INCVTE GENOMICS INC.  
 XX PA Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe LJ;  
 PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BM;  
 PI Thangaveilu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;  
 PI Hafalia AJA, Sanjanwala B, Marquis JP, Gorvad AE, Lee SY, Ison CH;  
 PI Baughin ME, Chawla NK, Nguyen DB, Swarnakar A, Zebartadian Y, Shah P;  
 PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE;  
 PI Burford N, Ramkumar J;  
 XX WPI: 2003-313243/30.  
 XX P-PSDB; ABJ96722.  
 XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
 PT treating and preventing diseases or conditions associated with the  
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
 PT infections.  
 XX Claim 5; Page 340-341; 345pp; English.  
 XX The invention describes a novel human isolated nucleic acid-associated  
 CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in  
 CC diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression or overexpression of NAAP, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
 CC disorders, or infections. These are also useful in assessing the effects  
 CC of exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of NAAP. The NAAP or its fragments are useful in screening  
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
 CC or in altering the expression of the target polynucleotide and compounds  
 CC that specifically bind to or modulate the activity of the polypeptide.  
 CC The microarray is useful in monitoring or measuring protein-protein  
 CC interactions, drug-target interactions, and gene expression profiles.  
 CC This sequence encodes a novel human nucleic acid-associated protein  
 CC (NAAP).  
 XX SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;  
 Query Match 7.8%; Score 96.8; DB 8; Length 1705;  
 Best Local Similarity 52.9%; Pred. No. 1.1e-09;  
 Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;  
 QY 7 GAAAGTGAACCAACCAATCGATATCATCTTATATATCTTCTTCTTCATCAGGCCCC 66  
 Db 695 GAATGTGCAAAACCAATTTTGTGATGCTTTCACACCTCACTCAATAAAGATCCAAACT 754  
 QY 67 AAAAAAGTATTTTGCATATGAAGGGTGTGATAAGCCTATAATCGACCATCATTTATTA 126  
 Db 755 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTAACTGCTCTCAACGCTT 814  
 QY 127 GAGCAACATTTAAGACCCACAGTAAGTATCGCGGTATTAATGTACAGTGGAGATGT 186  
 Db 815 AATAAACATAGAGAAATTCATCTCGAGAGAAACCTTACAAATGTAAA-----GAATGT 868

QY 187 GATAAGCATTTTTCAGAAATCAATTTGGAAACACATATTGTATCATATTCGAGAAAA 246  
 Db 869 GGCAGAGCCTTTTAAACAGACCTCACACCTTATTAGACATAAGAGAAATTCATCTAGAG 928  
 QY 247 AAACCATTCATTTTCAGTGTGTGTAAGGGTTAATTCCTCGAACACCTTTGAAAGA 306  
 Db 929 AAACCTCAATGTGAAGATGTGCAAGCCTTTAACCATCTATCGACCTTACTACA 988  
 QY 307 CATGAATCACCACATCAAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTT 366  
 Db 989 CATATATATTCATCTAGGGAAATTCCTTCAAAATGTGAGAAATGTGTTAGAGCTTTT 1048  
 QY 367 TATAACATCAATCTTTAGAGATCATATATTATCTGTT--CATGAAAAACATTAACG 423  
 Db 1049 AACCAAGCTCAAGCTTACTGACATAAGTTAATTCATACCGGAGAGAAATGCTACAA 1108  
 QY 424 TGTAAACAATGTAATAAGTTTTCATCTCGACCTTCAAAATAGCAACATTAATAA 483  
 Db 1109 TGTGAAGATGTGGCAAGTGTTTTAAACCGATCTTCAACCTTACAGACATAAAAAAAT 1168  
 QY 484 CATCATGGTG 493  
 Db 1169 CATACAGGAG 1178

RESULT 5  
 ACAS6456  
 ID ACAS6456 standard; cDNA; 2320 BP.  
 AC ACAS6456;  
 XX  
 XX  
 DT 06-JUN-2003 (first entry)  
 XX  
 DE Human signalling pathway polynucleotide probe SEQ ID NO 1054.  
 XX  
 XX Human; probe; ss; array element; Parkinson's disease;  
 KW signalling pathway population; cancer; adenocarcinoma; leukaemia;  
 KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.  
 XX  
 XX Homo sapiens.

XX US6500938-B1.  
 XX  
 PD 31-DEC-2002.  
 XX  
 XX 30-JAN-1998; 98US-00016434.  
 XX  
 XX 30-JAN-1998; 98US-00016434.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Au-Young J, Seilhamer JU;  
 XX WPI; 2003-352189/33.  
 XX  
 PT Combination of polynucleotide probes, useful as array elements in a  
 PT microarray for monitoring the expression of a number of target  
 PT polynucleotides.  
 XX  
 XX Claim 1; SEQ ID NO 1054; 65pp; English.

XX The invention relates to a combination which, comprises a number of  
 CC polynucleotide probes comprising a sequence selected from one of the 1490  
 CC sequences mentioned in the specification. The combination is useful as an  
 CC array element in a microarray for monitoring the expression of a number  
 CC of target polynucleotides. The microarray is particularly useful in the  
 CC diagnosis and treatment of cancer and immunopathology and neuropathology.  
 CC The microarray is useful in diagnostics and treatment regimens, drug  
 CC discovery and development, toxicological and carcinogenicity studies,  
 CC forensics and pharmacogenomics. The microarray is also useful for  
 CC monitoring progression of diseases and for developing sophisticated  
 CC profiles for the effects of currently available therapeutic drugs. The

CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
 CC and genomic fragments and in research and diagnostic applications. The  
 CC array can detect changes in expression in a large number of genes, coding  
 CC for different signalling pathway populations which can be used to diagnose  
 CC various diseases including cancer e.g. adenocarcinoma and leukaemia,  
 CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease  
 CC and Parkinson's disease. The present sequence represents a polynucleotide  
 CC probe of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docID=06500938B1  
 XX

SQL Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.6; DB 7; Length 2320;  
 Best Local Similarity 51.3%; Pred. No. 1.2e-09;  
 Matches 306; Conservative 0; Mismatches 279; Indels 12; Gaps 3;

QY 69 AAGATATATTGACATATGAAGGGTGTGATAAGCCTATAATCGACCATCATATTATGA 128  
 Db 964 AGAGAAACCCCTACAAATGTAAGAAATGTGTAAGCTTTTAACCGCATCTTCAACCTTAC 1023  
 QY 129 GCACATTTAAGACCCACACAGTAATGATCGACGTATAAATGTACAGTGACGATTGTGA 188  
 Db 1024 TACCATAGAAAAATTCATCTGAGAGAAACCTTACAAATGT-----GAAGATGTGG 1077  
 QY 189 TAAAGCATTTTTCAGAAAAATCACAATTTGGAAACACATATTGTATCATATTCCGAAAAA 248  
 Db 1078 CAAAGCCTTTTAAAGCAGTCTTCAAAACCTTACTACACATAAGATAATTCATCTGGAGAA 1137  
 QY 249 ACATTCCTCATTTGTCAGTGTGTGTAAGGGTGTATTTCTCGACACACCTTTGAAAGACA 308  
 Db 1138 ACCCTCAAAATGTAAAAAATGTGAAAGCCTTTTAACCGATCTGACACCTTACCACACA 1197  
 QY 309 TGAATCACCACATACAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAACGATTTA 368  
 Db 1198 TGAGGTAAATTCATCTGAGAGAAACCTTACAAATGTGAAAAATGTGAAAAAGCCTTTAA 1257  
 QY 369 TAAACA---TCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAACATTTAAGTG 425  
 Db 1258 TCATTTCTCACCTTACTACACATAAGATAATTCATCTGAGAGAAACCTTACAAATG 1317  
 QY 426 TAAACAATGTAATAAGTTTTCATCTCGACCTTCAAAATTTAGCACACATATAATTTAAACA 485  
 Db 1318 TAAAGATGTGTAAGCTTTTAAACACTCTTCAACCTTACTTAAACATAAGATAATTC 1377  
 QY 486 TCATGTGGATCTCTCTCTTATCATGTGATCATCTGTTGTTTAAAAATTTCCAAAC 545  
 Db 1378 TACTGGAGAGAGCCTTACAAATCTAAAGAAATGTGAAAAAGCCTTTAACCAATCCTCAA 1437  
 QY 546 TTGTCAGTATTACAAATTTCAATATAAAACAACTGCATCCAAACTTAAATGTCCTAAATG 605  
 Db 1438 ACTTACTGACATAAGAAAAATTCATCTGAGAGAAACCTTA---TGAATGTGAAAAATG 1494  
 QY 606 TGGTAAAGTTGTGTGCGAAAAAGGTTTATCTTCAATATGTTAAGTCATGATGA 662  
 Db 1495 TGGCAAGCTTTTAAACCGAGTCTCAATCTTACTAGACATAAGAAAAAGTCATACAGA 1551

## RESULT 6

AAS91317  
 ID AAS91317 standard; cDNA; 4563 BP.

XX AAS91317;  
 AC AAS91317;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX

DE DNA encoding novel human diagnostic protein #27121.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.

XX WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 PD 31-MAR-2000; 2000US-00540217.  
 XX 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX P-PSDB; ABG27130.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 1; SEQ ID NO 27121; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activities. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;  
 SQ

Query Match 7.7%; Score 95.2; DB 5; Length 4563;  
 Best Local Similarity 52.7%; Pred. No. 2.5e-09;  
 Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

QY 7 GAAGTGCAGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCATCAGTCCC 66  
 DB 1315 GAATGTGGCAAGCCCTTTAACAGTCTTACACCTTACTCAACATAAACCAATTCATACT 1374  
 QY 67 AAAAAATATATTTCACATATGAAGGGTGTGTATAAGCCTTAAATCGACCAATCATTTA 126  
 DB 1375 GGGGGAACCACTACAAATGTGAAGAAATGTGGCAAAAGCTTTTAAACCATCTTTCAGCCCTT 1434  
 QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACGTAATATGATGACAGTGGACGATTGT 186  
 DB 1435 ACTAAACATAAGATAATTATCATCTGGGGAGAAACCATACAAATGT-----GAAGAAATG 1488  
 QY 187 GATAAAGCATTTTTCAGAAAAATCACATTTGGAAAAACACATATTGTATCATCTTCGAAAAA 246  
 DB 1489 GCGAAGCTTTTAGGCAATCTTCACACCTTACTAGACATAAAGCAATTCATACCTGGAGAG 1548  
 QY 247 AACCAATTCATTTGTCAGTGTGTGTAAGGGTTAATTCGACACACACTTGAAGA 306  
 DB 1549 AAACCCCTACAAATGTGAAGAAATGTGGCAAAAGCTTTTAAACCAATTTCTCAGACCTTAGAAGA 1608  
 QY 307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAATTTGTCAAGAGCATTT 366

DB 1609 CATAAGATAATTCACTACTGGAAAGAACCCCTACAAATGTGAAGATGTGGAAAGCTTTT 1668  
 QY 367 TATAAACATCAATCTTTAAGA---CATCATATATTATCTGTTCATGTGAAATAACATTAACG 423  
 DB 1669 AGCCAGTCTCTCAACCTTTAGAAACCATCAGATAATTATCATCTGGAGAGAAACCCCTACAAA 1728  
 QY 424 TGTAAACAATGTAAATAAGTTTTTCACTGCACCTTCAAAATAGACACACATAAAATTTAAAA 483  
 DB 1729 TGTGAAGATGTGGTAAAGCTTTTAAAGTGGTGCATCAAAACCTTACTGTGATCAAGGTAATT 1788  
 QY 484 CATCATGGTG 493  
 DB 1789 CATACTGGAG 1798

RESULT 7  
 AAL51569  
 ID AAL51569 standard; DNA; 2597 BP.  
 XX  
 AC AAL51569;  
 XX  
 DT 10-APR-2003 (first entry)  
 XX  
 DE Human nucleic acid-associated protein coding sequence - SEQ ID No 52.  
 XX Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;  
 KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;  
 KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;  
 KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;  
 KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;  
 KW Crohn's disease; transgenic animal; animal model.  
 XX Homo sapiens.  
 OS  
 XX WO2003000864-A2.  
 PN  
 XX 03-JAN-2003.  
 PD  
 XX 20-JUN-2002; 2002WO-US021179.  
 XX  
 PF 22-JUN-2001; 2001US-0300518P.  
 PR 29-JUN-2001; 2001US-0301787P.  
 PR 29-JUN-2001; 2001US-0301792P.  
 PR 29-JUN-2001; 2001US-0301892P.  
 PR 29-JUN-2001; 2001US-0301893P.  
 PR 06-JUL-2001; 2001US-0303405P.  
 PR 06-JUL-2001; 2001US-0303442P.  
 PR 15-MAR-2002; 2002US-0364438P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 PA Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;  
 PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;  
 PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;  
 PI Berovsky ML, Yao MG, Walia NK, Bandman O, Lal PG, Becha SD, Lee SY;  
 PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y, Lu Y;  
 XX WPI; 2003-201420/19.  
 DR P-PSDB; AAO16419.  
 XX New nucleic acid-associated proteins and polynucleotides, useful for  
 PT diagnosing, treating or preventing cell proliferative (e.g. cancer),  
 PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune  
 PT disorders (e.g. AIDS).  
 XX Claim 12; Page 292-293; 312pp; English.  
 XX The invention comprises the amino acid and coding sequences of human  
 CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
 CC the invention are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of NAAP, such as: cell proliferative  
 CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis



AC ADB3624;  
 XX 04-DEC-2003 (first entry)  
 DT Human cDNA encoding clone THYM20071120.  
 DE Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
 KW tissue regeneration; cell regeneration; membrane protein;  
 KW signal transduction-related protein; transcription-related protein;  
 KW osteoporosis; neurological disease; cancer; tumour.  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 PH 170..1792  
 FT /tag= a  
 FT /product= "Clone THYM20071120 protein"  
 FT  
 XX  
 XX EP1308459-A2.  
 XX  
 XX 07-MAY-2003.  
 XX  
 XX 28-MAR-2002; 2002EP-00007401.  
 XX  
 XX 05-NOV-2001; 2001JP-00379298.  
 XX 25-JAN-2002; 2002US-0035978.  
 XX  
 XX (HELI-) HELIX RES INST.  
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 XX  
 XX WPI; 2003-450961/43.  
 XX P-PSDB; ADB55594.  
 XX  
 XX New polynucleotides and polypeptides, useful for developing a diagnostic  
 XX marker or medicines for regulation of their expression and activity, or  
 XX as targets of gene therapy.  
 XX  
 XX Claim 1; Page; 222pp; English.  
 XX  
 XX The invention discloses a polynucleotide comprising a sequence selected  
 XX from 1970 fully defined nucleotide sequences which encode novel  
 XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 XX or its partial peptide, an antibody binding to the polypeptide or peptide  
 XX of the polynucleotide, immunologically assaying the polypeptide or peptide  
 XX with the antibody of the polynucleotide by contacting the polypeptide or peptide  
 XX between the two, a transformant carrying the polynucleotide in an  
 XX expressible manner and an antisense polynucleotide. The oligonucleotide  
 XX is useful as a primer for synthesising the polynucleotide, or as a probe  
 XX for detecting the polynucleotide. The polynucleotides and encoded  
 XX proteins are useful as pharmaceutical agents and many disease-related  
 XX genes may be included in them, for developing a diagnostic marker or  
 XX medicines for regulation of their expression and activity, or as targets  
 XX of gene therapy. The genes are involved in tissue and/or cell  
 XX regeneration. Membrane proteins, signal transduction-related proteins,  
 XX transcription-related proteins, disease-related proteins and genes  
 XX encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 XX neurological diseases, cancer, tumours). The cDNA may be used to regulate  
 XX the activity or expression of the encoded protein to treat diseases. The  
 XX sequence presented is a cDNA of the invention. Note: Some of the sequence  
 XX data for this patent is not represented in the printed specification, but  
 XX is based on sequence information supplied by the European Patent Office.  
 XX  
 XX Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;  
 SQ  
 Query Match 7.4%; Score 91.8; DB 9; Length 2110;  
 Best Local Similarity 51.7%; Pred. No. fe-08;  
 Matches 314; Conservative 0; Mismatches 277; Indels 16; Gaps 4;  
 XX  
 QY 61 CGTCCAAAAGTATATTTGCACATATGAGGGTGTGATTAAGCCTATATCGACCATCA 120  
 DB 1253 CATACCGAAGAAACCCCTACAAATGTAAGAATGTGCAAGCTTTTAAACATCTCTCA 1312  
 QY 121 TTATTAGACCAACATTTAAGAACCCACAGTAATGATGACCGGTATTAATGTACAGTGGAC 180  
 DB 1313 GCCCTTACTACACATAAGAGAAATTCACACTGGAGAGAAACCCCTACAATGT-----GAA 1366  
 QY 181 GATTGTGATAAGCATTTTTTCAGAAAATACACATTTGGAAAACACATATTTGTATCATCTCC 240  
 DB 1367 GAATGTGGCAAGCTTTTAAACCGATCCTCAAACTTACTGTAACATAAGAACTTCATACT 1426  
 QY 241 GAAATAAACCAATTCATTTGTCTGTGTGTAAGGGTAAATTTCTCGACACACTTG 300  
 DB 1427 GGAAGAAACCCCTACAAATGTGAAGATGTGGCAAGCTTTTATCCCAATCCTCAAACTT 1486  
 QY 301 AAAAGACATGAATCAACCCATACAAAGTCATTTAAATGTATCATTTGAAAATTTGCAAGAA 360  
 DB 1487 ACTGAACATAAAAAAATTCATTTCTGGAGAGATACCTTACAAAGTGTGAAGATGTGGCAA 1546  
 QY 361 GCATTTTATAACATCAATCTTTA---AGACATCATATATATCTGTTCATGAATAACA 417  
 DB 1547 GCTTTTAAACACTCCTCATCTCCCTTACTACACATAAAGAAATTCATCTGGGAGAAACCC 1606  
 QY 418 TTAACGTGTAAACAATGTAATAAAGTTTTCACCTCGACCTTCAAAATTTAGCACACATAAAA 477  
 DB 1607 TACAAATGTGAAGATGTGGCAAGCTTTTAGCCGATCCTCAAACTTACTGTAACATAAG 1666  
 QY 478 TTAACAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCTCGTTGTTTAAAAA 536  
 DB 1667 ATAAATCTACTGGAGAGAAACCCCTATAAATGTGAGAGATGTGACAAAGCTTTTAAACAA 1726  
 QY 537 TTTCCAAACTTTGGTCAGTATTACAAATTTTATATATAAACAACACTGCATCCAAACTTAAATG 596  
 DB 1727 TCTGCAAACT-----TACTAAACATAAAAAAATACATCTGGAGAGAACTACAGAAC 1780  
 QY 597 TCCTAAATGTGTAAGCTTGTGTGGGAAAAAAGGTTTATCTTCACATATGTTAACTCA 656  
 DB 1781 TGAATGTGTGATTAATGATTTTGACAAACCTCAAACTTTTCTAAATATAAAGGAATCA 1840  
 QY 657 TGATGAT 663  
 DB 1841 TACTGGT 1847  
 RESULT 10  
 ADC5695  
 ID ADC5695 standard; cDNA; 2760 BP.  
 XX  
 XX AC ADC5695;  
 XX  
 XX DT 18-DEC-2003 (first entry)  
 XX  
 XX DE Human macroprotein-45-76 cDNA.  
 XX  
 XX KW human; macroprotein-45.76; dementia; arrhythmia; asthma; diabetes; ss;  
 XX gene.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX FH Key Location/Qualifiers  
 FT CDS 180..1430  
 FT /tag= a  
 FT /product= "Human macroprotein-45.76"  
 XX  
 XX PN CN1382725-A.  
 XX  
 XX PD 04-DEC-2002.  
 XX  
 XX PF 26-APR-2001; 2001CN-00112764.  
 XX  
 XX PR 26-APR-2001; 2001CN-00112764.  
 XX

(BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.  
XX Mao Y, Xie Y;  
XX WPI; 2003-269486/27.  
XX P-PSDB; ADCS6696.  
XX New human macroprotein-45.76, encoding polynucleotide, antagonist and  
XX recombinant production, useful for treating dementia, arrhythmia, asthma  
XX and diabetes.  
XX Claim 6; SEQ ID NO 1; 34pp; Chinese.  
XX  
XX The invention relates to a novel human macroprotein-45.76, the encoding  
XX polynucleotide, an antagonist and a method of recombinant production. The  
XX protein of the invention may be useful for treating dementia, arrhythmia,  
XX asthma and diabetes. The current sequence is that of the human  
XX macroprotein-45.76 cDNA of the invention.  
XX  
XX Sequence 2760 BP; 1016 A; 463 C; 499 G; 782 T; 0 U; 0 Other;  
SQ  
Query Match 7.4%; Score 91.8; DB 9; Length 2760;  
Best Local Similarity 52.1%; Pred. No. 1.1e-08;  
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;  
QY 11 GTGACGAAACCAATCGATATCATCTTTTAATATCTTCTTCTTCTTCATCAGTCCCAAA 70  
Db 712 GTGCGCAATCATTTTGCATGCTTTTCAAAATTAATCAACATAAGAAAATTCATCTAGAG 771  
QY 71 AGTATATTTGCACATATGAAGGTGTGATAAAGCCTTAAATCGACCATCATTTATAGAC 130  
Db 772 AGTATCTTCAAAATGTGAAGATGGTAAAGCCCTTAACTGGTCTCAACCCCTTACTA 831  
QY 131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGATCAGTGGACGATTTGATA 190  
Db 832 AACATAAGATAATTCATCTCGAGAAAACCCCTACAAATGT-----GAAGAATGTGGCA 885  
QY 191 AAGCATTTTTCAGAAATACATTTTGAACACATATTTGATCATCATTCGAAAAAACC 250  
Db 886 AAGCTTTTAAACCGGTCTCTCAATCTTACTTAAACATAAATAATTCATCTGAGAGAAAC 945  
QY 251 CATTCCATTTGTCAGTGTGTGTAAGGGGTAAATTTCTGACAAACACTTGAAGAACATG 310  
Db 946 CCTACAATGTGAAGATGTGGCAAGCTTTTAAACCGGTCTCAACCTTACTTAAACATA 1005  
QY 311 AATACCCCTACAAAGTCAATTTAATGATCAATTTGAAATTTGCAAGACATTTTATA 370  
Db 1006 AAGAATTCATACAGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGGCGCTTTAAC 1065  
QY 371 AACATCAATCTTT---AAGACATCATATATTTATCTGTTCATGAAAAACATTAACGTGA 427  
Db 1066 AGTCTCGATCTTAATAACATAAGAGATTCATATGGAAGATAAACCTTACAAATGTG 1125  
QY 428 AACAAATGATAAAGTTTTTCACTCGACCTTCAAAATAGACACACATAAATTAACATC 487  
Db 1126 AAGAATGTGGGAAAGCCCTTGAAGTATCTCAATCTTAAAAAACATAGATAATCCATA 1185  
QY 488 ATGGTGATCTCC 500  
Db 1186 CTGGGAAAAACC 1198  
RESULT 11  
ADB62468  
ID ADB62468 standard; cDNA; 3078 BP.  
XX ADB62468;  
AC ADB62468;  
XX  
DT 04-DEC-2003 (first entry)  
XX Human cDNA encoding clone FEBRA20063720.  
DE  
XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
KW

tissue regeneration; cell regeneration; membrane protein;  
KW signal transduction-related protein; transcription-related protein;  
KW osteoporosis; neurological disease; cancer; tumour.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 118..11878  
FT /\*tag= a  
FT /product= "Clone FEBRA20063720 protein"  
XX  
XX EPI308459-A2.  
XX  
XX 07-MAY-2003.  
XX 28-MAR-2002; 2002EP-00007401.  
XX 05-NOV-2001; 2001JP-00379298.  
XX 25-JAN-2002; 2002US-00350978.  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX WPI; 2003-450961/43.  
XX P-PSDB; ADB64438.  
XX  
XX New polynucleotides and polypeptides, useful for developing a diagnostic  
XX marker or medicines for regulation of their expression and activity, or  
XX as targets of gene therapy.  
XX  
XX Claim 1; Page; 222pp; English.  
XX  
XX The invention discloses a polynucleotide comprising a sequence selected  
XX from 1970 fully defined nucleotide sequences which encode novel  
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
XX or its partial peptide, an antibody binding to the polypeptide or peptide  
XX of the polynucleotide, immunologically assaying the polypeptide or  
XX peptide of the polynucleotide by contacting the polypeptide or peptide  
XX with the antibody of the encoded protein, and observing the binding  
XX between the two, a transformant carrying the polynucleotide in an  
XX expressible manner and an antisense polynucleotide. The oligonucleotide  
XX is useful as a primer for synthesising the polynucleotide, or as a probe  
XX for detecting the polynucleotide. The polynucleotides and encoded  
XX proteins are useful as pharmaceutical agents and many disease-related  
XX genes may be included in them, for developing a diagnostic marker or  
XX medicines for regulation of their expression and activity, or as targets  
XX of gene therapy. The genes are involved in tissue and/or cell  
XX regeneration. Membrane proteins, signal transduction-related proteins,  
XX transcription-related proteins, disease-related proteins and genes  
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,  
XX neurological diseases, cancer, tumours). The cDNA may be used to regulate  
XX the activity or expression of the encoded protein to treat diseases. The  
XX sequence presented is a cDNA of the invention. Note: Some of the sequence  
XX data for this patent is not represented in the printed specification, but  
XX is based on sequence information supplied by the European Patent Office.  
XX  
SQ Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;  
Query Match 7.4%; Score 91.8; DB 9; Length 3078;  
Best Local Similarity 52.1%; Pred. No. 1.1e-08;  
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;  
QY 11 GTGACGAAACCAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCAGTCCCAAA 70  
Db 650 GTGGCAATCATTTTGCATGCTTTTCAAAATTAACATAAGAAAATTCATCTAGAG 709  
QY 71 AGTATATTTGCACATATGAAGGTGTGATAAGCCCTAATATCGACCATCATTTATAGAC 130  
Db 710 AGTATCTTACAAATGTGAAGATGTGTAAGCCCTTTAACTGGTCTTCAACCCCTTACTA 769

QY 131 AACATTTAAGAACCCACAGTAATGATCGACCGGTATATAATGTATACAGTGGACGATTGTGATA 190  
 |||||  
 Db 770 AACATAGATAATTCATCTAGAGAGAAACCCCTACAAATGT-----GAAGAAATGTGGCA 823  
 QY 191 AAGCAATTTTCAGAAATACATTTTCGAAACACATATTTATATCAATTCGGAAGAAAAAAC 250  
 |||||  
 Db 824 AAGCTTTTACCCGGTCTCAATCTTTACTAAACATATAAATAATTCATCTAGAGAGAAAC 883  
 QY 251 CATTCATTTTTCAGTGTGTGTGTAAGGGGTAAATTCCTGCAACACATTTGAAAGACATG 310  
 |||||  
 Db 884 CCTACAAATGTGAAGATGTGGCAAGCTTTTAAACCGGTCTCAACCCCTTACTAAACATA 943  
 QY 311 AAATCACCCTACAAAGTCAATTTAAATGTACATTTTGAAATTTGTCAGAGCAATTTTATA 370  
 |||||  
 Db 944 AAGAATTCATACAGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCCTTTTAAAC 1003  
 QY 371 AACATCAATCTTT---AAGACATCATATATTTATCTGTTTCATGAAAAACATTAACGTGTA 427  
 |||||  
 Db 1004 AGTTCTCGATTCTTAATAAACATAAGAGAAATTCATATGGAAGATAAACCCCTACAAATGTG 1063  
 QY 428 AACAAATGATATAAGTTTTCACCTCGACCTTCAAAATAGCACAAACATTAATAAACATC 487  
 |||||  
 Db 1064 AAGAATGTGCAAGCCCTTAGAGTATCTCAATCTTAAACAAACATAAGATAATCCATA 1123  
 QY 488 ATGGTGGATCTCC 500  
 |||||  
 Db 1124 CTGGGAAAAAACC 1136

## RESULT 12

AAD55863  
 ID AAD55863 standard; cdna; 2026 BP.

XX AC AAD55863;

XX DT 07-AUG-2003 (first entry)

XX DE Human nucleic acid associated protein (NAAP)-34 cdna.  
 XX KW Human; nucleic acid associated protein; NAAP; stroke; AIDS; nontropic;  
 KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;  
 KW developmental disorder; antiinflammatory; neuroprotective; thymomimetic;  
 KW Cushing's syndrome; infection; gene therapy; cytostatic; anticonvulsant;  
 KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;  
 KW gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 126..1973  
 FT /\*tag= a  
 FT /product= "Human NAAP protein"

XX PN W02003006618-A2.

XX PD 23-JAN-2003.

XX PF 10-JUL-2002; 2002W0-US021971.

XX PR 12-JUL-2001; 2001US-0305089P.

XX PR 12-JUL-2001; 2001US-0305104P.

XX PR 13-JUL-2001; 2001US-0305325P.

XX PR 13-JUL-2001; 2001US-0305390P.

XX PR 18-JUL-2001; 2001US-0306960P.

XX PR 20-JUL-2001; 2001US-0306694P.

XX PR 27-JUL-2001; 2001US-0308170P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;

XX PI Yue H, Baughn MR, Emerling BK, Lal PG, Lu DM, Forsythe LJ, EA;

XX PI Ramkumar J, Li JX, Becha SD, Duggan SW, Sanjanwala MW, Lee E;

PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG;  
 PI Barroso I, Tran B, Walia NK, Hafalia AJA, Nguyen DB, Lu Y;  
 PI Arvizu CS;  
 XX WPI; 2003-221732/21.  
 DR P-PSDB; AAB37047.

XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
 PT treating and preventing diseases or conditions associated with the  
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
 PT infections.

XX Claim 5; Page 259-260; 260pp; English.

XX The invention relates to human nucleic acid associated proteins (NAAP)  
 and their corresponding nucleic acid sequences. The invention is useful  
 CC in diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression or overexpression of NAAP, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
 CC disorders, or infections. It is also useful in assessing the effects of  
 CC exogenous compounds on the expression of nucleic acid and amino acid  
 CC compounds for which acts as their agonist or antagonist. The microarray  
 CC is useful in monitoring or measuring protein-protein interactions, drug-  
 CC target interactions, and gene expression profiles. NAAP DNA is used in  
 CC gene therapy. The present sequence is human NAAP cdna

XX Sequence 2026 BP; 599 A; 422 C; 396 G; 509 T; 0 U; 0 Other;

Query Match 7.3%; Score 90.4; DB 7; Length 2026;  
 Best Local Similarity 51.8%; Pred. No. 2e-08;  
 Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

QY 7 GAAAGTGACGAAACCAAAATCGATATCATCTTTTAATATCTCTTCTTCTCATCGTCCC 66  
 |||||  
 Db 1002 GAATGTGGCAAGCCTTCTACCATTTCTCACACCTTACTACATAGGTAATCTACT 1061  
 QY 67 AAAAGTATATTGACATATGGAAGGGTGTGATGAAGCCTTATATCGACCATCATTTATTA 126  
 |||||  
 Db 1062 GGAGAGAGACCCCTTCAAAATGTGAAGAATGTGTAAAGCTTTTAAACCCCTTCAGCCCT 1121  
 QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186  
 |||||  
 Db 1122 ACTACACATAGTTCATTCATGTTTAAGAAAAACCCCTACAAATGT-----GAAGATGT 1175  
 QY 187 GATAAGCATTTTTCAGAAAAATCAGATTTGGAACACATATTTGATTCACATTCGAAAAA 246  
 |||||  
 Db 1176 GACAAAGCTTTTAAACCGATTCTCATCCTTACTATAACATAAGATAATTCATTCTGGAG 1235  
 QY 247 AAACCATTTCCATTGTTCAGTGTGTGTAAGGGGTTAATTTCTGACAAACACTTGAAGA 306  
 |||||  
 Db 1236 AAATCTTACAAATGTGAACATGTGGCAAGGCTTTAACTGGTCTTCAACCCCTACAAA 1295  
 QY 307 CATGAATCACCACATACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGACATTT 366  
 |||||  
 Db 1296 CATAGAAGATTCATCTACTGGAGAGAAACCCCTACAAATGTGAAGATGTGCAAGCCCTT 1355  
 QY 367 TATAAACAT---CAATCTTTAAGACATCATATATTTCTGTTCATGAAAAACATTAACG 423  
 |||||  
 Db 1356 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATCTACTGGAGAGAAACCTACAA 1415  
 QY 424 TGTAAACATGTAATAAGTTTTCACCTCGACCTTCAAAATTTAGCACACATATTAATAA 483  
 |||||  
 Db 1416 TGTGAAGATGTGGCAAGCCCTTTAAACACTCTCTCAAACTTACTATACATATAGATTA 1475  
 QY 484 CATCATGGTGGATCTCCT 501  
 |||||  
 Db 1476 CATACTGGAGAGAAACCT 1493

## RESULT 13

Db	1318	CATAGAAGATTCATACACTGGAGAGAAACCCATCAAAATGTGAAGAATGTGGCAAGCCTTT	1377
Qy	367	TATAAACAT---CAATCTTTTAAGACATCATATATTATCTGTTTCATGAAAAAACAATTAACG	423
Db	1378	AATGTGTCITCACACCTTACTACACATAAGATGATTCATACTGGAGAAACCCATCAAA	1437
Qy	424	TGTAACAATGTAAATAAGTTTTCACCTCGACCTTCAAAATTAGCACACAATAAATTAATA	483
Db	1438	TGGAAGAATGTGGCAAGCCTTTAAACCACTCCTCAAACTTACTATACATAAGATAATT	1497
Qy	484	CATCATGGTGGATCTCCT	501
Db	1498	CATACTGGAGAGAAACCT	1515
RESULT 14			
ID	ADA52931	standard; cdna; 2230 BP.	
AC	XX		
AC	ADA52931;		
DT	20-NOV-2003	(first entry)	
XX			
DE	Human coding sequence, SEQ ID 499.		
XX			
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;		
KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;		
KW	inflammatory disease; osteoporosis; neurological disease; gene; ss.		
OS	Homo sapiens.		
XX			
PN	EP1293569-A2.		
XX			
PD	19-MAR-2003.		
XX			
PF	21-MAR-2002; 2002EP-00006586.		
XX			
PR	14-SEP-2001; 2001JP-00328381.		
PR	24-JAN-2002; 2002US-0350435P.		
XX			
PA	(HELI-) HELIX RES INST.		
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.		
XX			
PI	Iscgal T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;		
PI	Yamanoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;		
PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;		
DR	WPI: 2003-395539/38.		
DR	P-PSDB; ADA54570.		
XX			
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory		
PT	and/or membrane proteins, useful for developing medicines for diseases in		
PT	which the gene is involved, or as target molecules for gene therapy.		
XX			
PS	Claim 1; SEQ ID NO 499; 205pp; English.		
XX			
CC	The present invention relates to novel human secretory or membrane		
CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-		
CC	ADA54071). The coding sequences are useful in the gene therapy of		
CC	diseases caused by abnormalities of the proteins, e.g. cancer,		
CC	inflammatory diseases, osteoporosis or neurological disease.		
XX			
SQ	Sequence 2230 BP; 831 A; 434 C; 400 G; 565 T; 0 U; 0 Other;		
Query Match 7.3%; Score 90.2; DB 7; Length 2230;			
Best Local Similarity 50.4%; Pred. No. 2.2e-08;			
Matches 332; Conservative 0; Mismatches 313; Indels 14; Gaps 4;			
Qy	7	GAAAGTGACGAACCAAAATCGATATCATCTTTAAATATCTTCTTCTTCATCACGTCCC	66
Db	656	GAATGTGACAAATCACTTTGCGATGCTTTTCACGCTTAATCAACATAAAAAATTCATCT	715

QY 67 AAAAGATATATTCACATATGAAGGGTGTGATGAAGCCCTATATCGACCATCATTTA 126  
 Db 716 AGAGGAAATTTCTACAAATGTGAGAGGTGTGGAAAACCTTTAACTGGTCCAAACCTT 775  
 QY 127 GAGCAACATTTAAGAACCCAGTAATGATCGACCGGTATTAATGTACAGTGGACGATTGT 186  
 Db 776 TCTAARCCCTAAGAAAATTTCTACTGGAGAAAAACCCCTACAAAATGTGAAGTA-----TGT 829  
 QY 187 GATAAAGCATTTTTCAGAAAATCACATTTGGAACACATATTTATCACATTCGGAATA 246  
 Db 830 GGAAAAGCCTTTCACCAATCCTCAATCCTTACTAACAATAAGTAATTCGTACTGAGNA 889  
 QY 247 AAACCATTCATTTGTCAGTGTGTGGTAAAGGGGTAAATTTCTGCAACACACTTGAAGA 306  
 Db 850 AAACCCCTATAATGTGACACTGTGGGAAAGCCCTTTAAACAGCTCTTACACCTTACTAGA 949  
 QY 307 CATGAATCACCCATACAAAGTCAATTTAAATGTACATTTGMAAATGTGCAAGCATTT 366  
 Db 950 CATAGATATTCATCTAGAGAAACCTTACAAATGTGAACATGTGGCAAGTCTTT 1009  
 QY 367 TATAAACATCATC---TTTAAAGACATCATATTTATCTGTCATGAAAAAATTAACG 423  
 Db 1010 AAGCAGTCCCCAACCCCTTACTAACAATCATGATAATTTACTGGAGAGAACCATACAAA 1069  
 QY 424 TGTAACAATGTATAAAGTTTTCACGACCTTCAAAATTTAGCAACATATAATTAATA 483  
 Db 1070 TGTGAGGAATGTGCAAGCTTTTACCTATCTTACACTTACTGACATAGAAATTT 1129  
 QY 484 CATCATGGTG-GATCTCCTCTTATCAATGTGATCATCTGTTGTTTAAATAATTCOA 542  
 Db 1130 TACACTAGAGAGAAAGCTTCAAAATGTGAAGATGTGGCAAGCCCTTTAACCAAGTTTCA 1189  
 QY 543 AACTTGTGTCAGTATTACAAATTTCAATAAACAACCTGCATCCAAAACCTTAAATGCTTAA 602  
 Db 1190 ACCCTTATACACATAGATAATTCATAGCAGAGAAACCC---CACAAATGTGAAGA 1245  
 QY 603 ATGTGTTAAGTTTGTGTTGGGAAAAAGTTTATCTTCAATATGTTAAGTCATGATG 661  
 Db 1246 ATGTGGCAGAGCTTTTAAACCAAGTCCGCAAGCTCACTGAACATAAGTTAATTCATAC 1304

## RESULT 15

ADC30762

ID ADC30762 standard; cDNA; 2064 BP.

XX ADC30762;

XX 18-DEC-2003 (first entry)

XX Human novel cDNA sequence, SEQ ID NO:844.

XX Human; diagnostic; drug screening; forensics; gene mapping;  
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KW ulcers; osteoporosis; autoimmune disease; cancer;  
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;  
 KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
 KW gene therapy; chromosome 19q12; gene; ss.

XX Homo sapiens.

XX WO2003029271-A2.

XX 10-APR-2003.

XX 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSE-) HYSEQ INC.

XX Tang Ty, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
 PI

Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 Haley-Vicente D, Drmanac RT;

WPI; 2003-371981/35.

P-PSDB; ADC31733.

New polynucleotide and polypeptide useful for diagnosing, preventing or  
 treating conditions such as neurodegenerative diseases, anemias, platelet  
 disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 cancer.

Claim 1; SEQ ID NO 844; 1185pp; English.

The invention relates to 971 novel human cDNA sequences (ADC29919-  
 ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
 invention also relates to nucleic acid sequences over 99% identical with  
 the novel human cDNAs. The invention additionally encompasses expression  
 vectors and host cells comprising a nucleic acid of the invention; the  
 recombinant production of a polypeptide of the invention; an antibody  
 against a polypeptide of the invention; a method of detecting  
 polynucleotides or polypeptides of the invention; and methods of  
 identifying a compound which binds to a polypeptide of the invention.  
 The invention further discloses methods of preventing, treating or  
 ameliorating a medical condition; kits comprising polynucleotide probes  
 and/or monoclonal antibodies for carrying out the methods of the  
 invention; methods for the identification of compounds that modulate the  
 expression or activity of the polynucleotide and/or polypeptide; and 767  
 contig sequences corresponding to the cDNA sequences of the invention  
 (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
 -ADC33394). The nucleic acids and polypeptides of the invention are  
 useful in diagnostic, drug screening, forensics, gene mapping, in the  
 identification of mutations responsible for genetic disorders or other  
 traits, for assessing biodiversity, and in producing many other types of  
 data and products dependent on DNA and amino acid sequences. They are  
 also used for treating diseases such as Parkinson's disease, Alzheimer's  
 disease and other neurodegenerative diseases, anaemia, platelet  
 disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 cancer. The nucleic acids may also be used as hybridisation probes or  
 primers, and in the recombinant production of a protein. The polypeptides  
 are also useful in regenerating antibodies, as molecular weight markers,  
 and as food supplements. The present sequence represents a specifically  
 claimed human cDNA sequence of the invention. Note: The sequence data for  
 this patent did not form part of the invention. Note: The sequence data for  
 obtained in electronic format directly from the printed specification, but was  
 ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 2064 BP; 703 A; 375 C; 449 G; 537 T; 0 U; 0 Other;

## Query Match

Best Local Similarity 7.2%; Score 89.2; DB 9; Length 2064;

Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

QY 69 AAAGTATATTTCCACATATGAAGGGTGTGATAAAGCCCTATATCGACCATCATTTAGA 128  
 Db 828 AGAGAAACCTTACAAAATGTGAAGATGTGACAAAAGTTTGTAGTCGTAATCACACCTTGA 887  
 QY 129 GCAACATTTAAGAACCCACAGTAAATGATGCGACCGCTATAAATGTACAGTGACCATTTGA 188  
 Db 888 AAGACATAGGAGAAATTCATCTGGAGAGAAACCGTACAAATGTAAGGT-----TTGTGA 941  
 QY 189 TAAAGCATTTTTCAGAAAATCACATTTGGAAAACACATATTTGTATCATCATCCGAAAAA 248  
 Db 942 CAAGGCTTTCAGACGCTGATTTCACACCTGGCAACATATTTGTAATTCACACTGGAGAA 1001  
 QY 249 ACCATTCATTTTCAGTGTGTGTTAAGGGGTAAATTTCTCGAACACACTTTGAAAAAGACA 308  
 Db 1002 ACCTTACAGTGAATGATGTGGCAAGACCTTTGTTTCAAAATTCATCTCTTGTATGCA 1061  
 QY 309 TGAATTCACCCATAC---AAAGTCATTTAATGTACATTTGAAATTTGTCAAGAGCATT 365  
 Db 1062 TAAAGTCATTCATCTGAGAGAAATGTACAAAGTGAATGTGCGAAGGTTTTTAA 1121  
 QY 366 TTATAAACATCAATCTTTAAGACATCATATATTATCTGTTTCATGAAAAAACATTAACGTG 425

Db	1122	TCACAAATCAAACCTTGCATGTGCATCATATAGACTTCATCTGGAGAGAAACCTTACAAGTG	1181
Qy	426	TAAACAATGTAATAAAGTTTTTCACTCGACCTTCAAAAATTAGCACAAACATATAAATTAAACA	485
Db	1182	TAATGATGTGGCARGGTTTTTAATTGAAATCAAACCTTGAACATCATCTAGAGTTCA	1241
Qy	486	TCATGGTGATCTCCT	501
Db	1242	TATCGAGAGAAACCT	1257

Search completed: May 9, 2004, 06:07:07  
Job time : 549.661 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 05:31:45 ; Search time 3619.38 Seconds  
(without alignments)  
10222.531 Million cell updates/sec

Title: US-09-831-804-2

Perfect score: 1239

Sequence: 1 atgagtgaagtgcacgaac.....catcagtgattcttcgataa 1239

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_estba:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	182.2	14.7	628	BZ298000	BZ298000 CG3957.f1
C 2	167.6	13.5	959	29	AL405671 T7 end of
C 3	98.2	7.9	1609	11	AK032220 Mus muscu
4	97.8	7.9	2647	11	BC047646 Homo sapi

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	97.6	7.9	745	9	AU123448
6	92.4	7.5	2672	11	BC036394
7	92	7.4	591	28	AQ005136
8	91	7.3	533	28	B93387
9	91	7.3	1163	13	EX414627
10	90.4	7.3	2694	11	BC022527
11	90.4	7.3	2698	11	BC037782
12	90.2	7.3	2253	11	BC028252
13	89.2	7.2	604	10	BE161630
14	89.2	7.2	3609	11	BC032590
15	88.6	7.2	535	9	AL705393
16	88.6	7.2	2174	11	AK033001
17	88.6	7.2	2330	11	BC037426
18	88.6	7.2	2622	11	AK033958
19	88.4	7.1	1084	13	EX456765
20	88	7.1	895	13	BU195416
21	87.6	7.1	832	13	EX437291
22	87.6	7.1	1201	13	EX355654
23	87.4	7.1	1200	13	EX437758
24	87.2	7.0	579	28	AQ47265
25	87.2	7.0	687	28	AQ389266
26	87.2	7.0	907	13	BQ423752
27	87	7.0	660	14	CD770140
28	87	7.0	842	13	BU507408
29	86.8	7.0	521	28	AQ475826
30	86.8	7.0	705	28	AQ194882
31	86.8	7.0	860	28	AQ749175
32	86.8	7.0	897	13	BQ431141
33	86.8	7.0	1906	11	AK013043
34	86.6	7.0	694	12	BG596505
35	86.4	7.0	874	13	BU171453
36	86.4	7.0	2791	11	BC020045
37	86.2	7.0	523	14	CA874049
38	86	6.9	557	10	BF817669
39	86	6.9	667	14	CF135920
40	86	6.9	685	29	AG149704
41	86	6.9	801	14	CD657347
42	85.6	6.9	711	14	CB154257
43	85.6	6.9	712	9	AI792344
44	85.4	6.9	682	12	BG818171
45	85.4	6.9	2106	11	AK030776

## ALIGNMENTS

RESULT 1 BZ298000/c 628 bp DNA linear GSS 31-OCT-2002  
LOCUS CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata  
DEFINITION genomic clone CG3957, genomic survey sequence.

ACCESSION BZ298000  
VERSION BZ298000.1 GI:24440936

KEYWORDS GSS.

SOURCE Candida glabrata

ORGANISM Candida glabrata

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 (bases 1 to 628)

Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.

Evidence from comparative genomics for a complete sexual cycle in

the 'asexual' pathogenic yeast Candida glabrata

Genome Biol. 4 (2), R10 (2003)

JOURNAL MEDLINE

22508158

PUBMED 12620120

COMMENT Contact: Wong S

Department of Genetics, Smurfit Institute

Trinity College Dublin

Dublin 2, Ireland

Tel: 353 1 6082319

Fax: 353 1 6798558

Email: swong@tcd.ie

Class: plasmid ends.

```
FEATURES
  source
    Location/Qualifiers
      1..628
        /organism="Candida glabrata"
        /mol_type="genomic DNA"
        /strain="CBS 138"
        /db_xref="taxon:5478"
        /clone="CG3957"
        /clone_lib="Candida glabrata Random Genomic Library"

ORIGIN
  Query Match      14.7%; Score 182.2; DB 28; Length 628;
  Best Local Similarity 59.5%; Pred. No. 3.2e-24;
  Matches 348; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

Qy 84 ATATGAAGGCTGATGAAGGCTTAATCGACCATCATTTATAGAGCAACATTTAGAAC 143
Db 625 ATACGATTAATGTGACAAAGGCTTTACAGGCCCTCTTGTCTACAGAACACAGATAC 566
Qy 144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATGTGATAAGACATTTTCAG 203
Db 565 ---AGTGCATCTGGAGGAAACCTTGGAAATGTAATCAATGTAAGAGTTCAITTAATA 509
Qy 204 AAATCATCATTTGGAACACATATGTATCATCTCGAAAGACATGAAATCCACCCATAC 263
Db 508 AAAGATCCACTTAGAGAGACATTTATACACACACAGATGAAGACCGTTTATTGTTTC 449
Qy 264 AGTGTGCTGAAGGGGTTAATCTCGACACACTTGAAGAGACATGAAATCCACCCATAC 323
Db 448 ATTTGTGGGAAGGGCTAATTACTAGGCAACACTGAAGACATCAGGTAACTCAGAC 389
Qy 324 AAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCAATTTTATAACATCATCTTT 383
Db 388 CAAATCGTTCAATTTGTAATGAAGGTTGTAATGAGAGTTTTCACAGCACCCCTCAAT 329
Qy 384 AAGACATCATATATTCTGTTTCATGAAAAACATTAACGTGTAAACAAATGTAATAAGT 443
Db 328 AAGGCTCATATTTAGCAGTTTCAATTACAAGTCTAAATGTCAAGATGCACAAAG 269
Qy 444 TTCTACTCGACCTTCAAAATTAGCACACATAAATAAATCAATCAT-----GGTGATC 497
Db 268 TTTTCAAAGACCTTACAGACTCAAAATTCATAGCCAAACACACCATATCCAGATGTTGT 209
Qy 498 TCCTGCTTATCAATGTGATCATCTCGTGTGTTTAAATAATTTCCAACTTGGTCAGTATT 557
Db 208 CAATGCTTATCAATGATCCTTCAGTGTGTTGTTTCAAGTTCAAACATGGTCGCTT 149
Qy 558 ACAATTTTCATATAAAACAACTGCATCCAAAACCTTAAATGTCTTAAATGTGTAAGGTTG 617
Db 148 AAGATTACATGTTAAATAATGATCATCCGAAGTAAATGCGCTATATGCAATAAACCTG 89
Qy 618 TGTGGGAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db 88 TGTGGGGAAGACGGTTTAAATATGCAATGAAATAATCCAGATGA 44

RESULT 2
CNS06MMP/c 959 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AU00AA005F10 of library AU00A from strain CBS 3082
DEFINITION of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL405671
VERSION AL405671.1 GI:12168715
KEYWORDS GSS
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri
REFERENCE 1 (bases 1 to 959)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Monigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,

Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 959)
Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Gaillardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
FEBS Lett. 487 (1), 56-60 (2000)
20584719
11152884
3 (bases 1 to 959)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segrifgenoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
  source
    Location/Qualifiers
      1..959
        /organism="Saccharomyces kluyveri"
        /mol_type="genomic DNA"
        /strain="CBS 3082"
        /db_xref="taxon:4934"
        /clone="AU00AA005F10"
        /clone_lib="AU00A"
        /note="end : 97"
        /note="complement (<71..>958)"
        /note="Similar to Saccharomyces cerevisiae ORF YPR186c [
          PZFI; TFIIF (transcription initiation factor) ]"
        /evidence=not_experimental

  misc_feature
    complement (<71..>958)
    /note="Similar to Saccharomyces cerevisiae ORF YPR186c [
      PZFI; TFIIF (transcription initiation factor) ]"
    /evidence=not_experimental

ORIGIN
  Query Match      13.5%; Score 167.6; DB 29; Length 959;
  Best Local Similarity 57.2%; Pred. No. 1.6e-21;
  Matches 324; Conservative 1; Mismatches 235; Indels 6; Gaps 1;

Qy 211 CATTTGGAAACACATATTGTATCACATTCGAAAAAACCATTCCATTGTTCAAGTGTGT 270
Db 958 CACTTGGAAAGACATATGTTTCACGCATTCGATAATAAACCATTTTCATTGTCCTATTGT 899
Qy 271 GGTAAAGGGGTTAATTCGACAACTTGAAGACATGAAATCACCACATCAAAAGTCA 330
Db 898 GGTAAAGGTTTACTACAAAGCAGCGATTAAAGACATGAGATCACATACCAATCG 839
Qy 331 TTTAAATGTACATTTGAAATTTGCAAGAAGCATTTTATAAACATCAATCTTTAAGACAT 390
Db 838 TTCAAATGTGAATGAGGGCTGCAAGCAATGTTTTCACAGCATTTCTCAACTAAGGTCA 779
Qy 391 CATATATTAATCTGTTTCATGAAAAAACATTAACGTGTAAACATGTAATAAGCTTTTCACT 450
Db 778 CATACACTTTTCGGTACACTTGCAGAAATTCAGCTGTGAGCATTTGGGTAATAATGTTCCA 719
Qy 451 CGACCTTTCAAAATTAGCACAAATATAATAAACATCATGTTGGATCTCTCTGCT----- 504
Db 718 AGGCCATATCGTTCGAAACCATCTTTCGAAAGCATCACAATGTGATGTTTGAAACAAA 659
Qy 505 TATCAATGTGATCATCTCGTGTGTTTTAAAAATTTCCAAAACCTTGGTCAGTATTAACAATTT 564
Db 658 TATCAATGTACTTACACGCGGTGCACAGAGACTTTTAAAAACCTGAGCTCTCTCTCAGCAG 599
```

QY 565 CATATAAACAACATGCGATCCAAAACCTTAATAGTCCTTAATATGCTGTAAGGTTGTGTGGG 624  
 Db 598 CATATCAAGCGGACCATCCCAAGTTCACCTGTAACTGTATGTGCGAAAGCTTGTGTGGT 639  
 QY 625 AAAAAAGGTTTATCTTACATATGTTTAAAGTCATGATCTTACCATGATCAAAATATGG 684  
 Db 538 GAGTCTGCTACAGATGCACATCGCATCGAGTCTCATGACGAGCTTTAGTATATAAAATGG 479  
 QY 685 ACTTGTGATTATGTTGATGGGGAATTTGCCAAGAAATGAATTAAGTTGTAACATTAT 744  
 Db 478 AAATGTACCATTTTGTACGCAAGTCTCTTTGCTAAGAAGCTGATCTTTTATCTCACTAC 419  
 QY 745 AATATCTTCATGATGTAATATCC 770  
 Db 418 ATGGAACATCATAGGATGATATACC 393

RESULT 3  
 AK032220  
 LOCUS  
 DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
 enriched library, clone:6430503015 product:MSZF33 (FRAGMENT)  
 homolog [Mus musculus], full insert sequence.  
 ACCESSION AK032220  
 VERSION AK032220.1 GI:26328050  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 1 Carninci, P. and Hayashizaki, Y.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 542  
 543  
 544  
 545  
 546  
 547  
 548  
 549  
 550  
 551  
 552  
 553  
 554  
 555  
 556  
 557  
 558  
 559  
 560  
 561  
 562  
 563  
 564  
 565  
 566  
 567  
 568  
 569  
 570  
 571  
 572  
 573  
 574  
 575  
 576  
 577  
 578  
 579  
 580  
 581  
 582  
 583  
 584  
 585  
 586  
 587  
 588  
 589  
 590  
 591  
 592  
 593  
 594  
 595  
 596  
 597  
 598  
 599  
 600  
 601  
 602  
 603  
 604  
 605  
 606  
 607  
 608  
 609  
 610  
 611  
 612  
 613  
 614  
 615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623  
 624  
 625  
 626  
 627  
 628  
 629  
 630  
 631  
 632  
 633  
 634  
 635  
 636  
 637  
 638  
 639  
 640  
 641  
 642  
 643  
 644  
 645  
 646  
 647  
 648  
 649  
 650  
 651  
 652  
 653  
 654  
 655  
 656  
 657  
 658  
 659  
 660  
 661  
 662  
 663  
 664  
 665  
 666  
 667  
 668  
 669  
 670  
 671  
 672  
 673  
 674  
 675  
 676  
 677  
 678  
 679  
 680  
 681  
 682  
 683  
 684  
 685  
 686  
 687  
 688  
 689  
 690  
 691  
 692  
 693  
 694  
 695  
 696  
 697  
 698  
 699  
 700  
 701  
 702  
 703  
 704  
 705  
 706  
 707  
 708  
 709  
 710  
 711  
 712  
 713  
 714  
 715  
 716  
 717  
 718  
 719  
 720  
 721  
 722  
 723  
 724  
 725  
 726  
 727  
 728  
 729  
 730  
 731  
 732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754  
 755  
 756  
 757  
 758  
 759  
 760  
 761  
 762  
 763  
 764  
 765  
 766  
 767  
 768  
 769  
 770  
 771  
 772  
 773  
 774  
 775  
 776  
 777  
 778  
 779  
 780  
 781  
 782  
 783  
 784  
 785  
 786  
 787  
 788  
 789  
 790  
 791  
 792  
 793  
 794  
 795  
 796  
 797  
 798  
 799  
 800  
 801  
 802  
 803  
 804  
 805  
 806  
 807  
 808  
 809  
 810  
 811  
 812  
 813  
 814  
 815  
 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832  
 833  
 834  
 835  
 836  
 837  
 838  
 839  
 840  
 841  
 842  
 843  
 844  
 845  
 846  
 847  
 848  
 849  
 850  
 851  
 852  
 853  
 854  
 855  
 856  
 857  
 858  
 859  
 860  
 861  
 862  
 863  
 864  
 865  
 866  
 867  
 868  
 869  
 870  
 871  
 872  
 873  
 874  
 875  
 876  
 877  
 878  
 879  
 880  
 881  
 882  
 883  
 884  
 885  
 886  
 887  
 888  
 889  
 890  
 891  
 892  
 893  
 894  
 895  
 896  
 897  
 898  
 899  
 900  
 901  
 902  
 903  
 904  
 905  
 906  
 907  
 908  
 909  
 910  
 911  
 912  
 913  
 914  
 915  
 916  
 917  
 918  
 919  
 920  
 921  
 922  
 923  
 924  
 925  
 926  
 927  
 928  
 929  
 930  
 931  
 932  
 933  
 934  
 935  
 936  
 937  
 938  
 939  
 940  
 941  
 942  
 943  
 944  
 945  
 946  
 947  
 948  
 949  
 950  
 951  
 952  
 953  
 954  
 955  
 956  
 957  
 958  
 959  
 960  
 961  
 962  
 963  
 964  
 965  
 966  
 967  
 968  
 969  
 970  
 971  
 972  
 973  
 974  
 975  
 976  
 977  
 978  
 979  
 980  
 981  
 982  
 983  
 984  
 985  
 986  
 987  
 988  
 989  
 990  
 991  
 992  
 993  
 994  
 995  
 996  
 997  
 998  
 999  
 1000

Katoch, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, K., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/  
 Location/Qualifiers  
 1. 1609  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM:DB:6430503015"  
 /db\_xref="MG1:2395676"  
 /db\_xref="taxon:10090"  
 /clone="6430503015"  
 /sex="male"  
 /tissue\_type="olfactory brain"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 330. 1223  
 /note="unnamed protein product; MSZF33 (FRAGMENT) homolog [Mus musculus] (SPTR|O88251, evidence: FASTY, 87.1%ID, 100%length, match=255)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC27766.1"  
 /db\_xref="GI:26328051"  
 /translation="MLETYRNNAVGVNWDSENIIEHSSRRRGRHNRHNTGKPYE  
 GIOYGEAPVHSSLQMRKIHTGKRYKNCQDKAYSRHSILQIHKTHSGKEPYECN  
 OCGKAFTHSHLKHVHTTCEKPKYKNCQDKAFACHNKLQKHRTHTGKPYKDCQ  
 NKAIVFESYLQVHKHTHTGKPKYKNCQDKAFARSHLKVHKIHTHTGKPYKNCQCK  
 ALAVHSILQVQRHTHTGKPKYKNCQDKAFANQSYFQVHKIHTHTGKPYKDCQCKAF  
 VGSSDLKKHVRHTGRETQW"  
 polyA\_signal  
 1594. 1599  
 /note="putative"  
 polyA\_site  
 1609  
 /note="putative"

ORIGIN  
 Query Match 7.9%; Score 98.2; DB 11; Length 1609;  
 Best Local Similarity 51.2%; Pred. No. 1.5e-08;  
 Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;  
 QY 60 ACCTCCCAAAAGTATATTTCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119  
 Db 521 ACATCTCTGGAGAAACCGCTACAATGCAATGCAATGATGATTAAGCCTATTCCAGACACAG 580  
 QY 120 ATTATTAGAGCAACATTTTAAGAACCCACAGTAATGATCGACCGTATTAATGATGAGTGA 179  
 Db 581 CATCTTCAAAATACATAAAAGAACACATAGTGGAGAGAAACCGCTATGAATGTA-----A 634  
 QY 180 CGATTGTGATGAAGCATTTTTCAGAAATTCATTTGGAACACATATTGTTATCATCATTC 239  
 Db 635 TCAATGTGGTGAAGCCTTTTACAAACACAGTCATCTCAAAATACATATGTTTACACATAC 694

QY 240 CGAAGAAAAACCACTTCATTTGTCAGTGTGTGTAAAGGGTTAAATCTCGACAAACACTT 299  
 Db |||||  
 695 TGGAGAGAAACCTTACAAATGTAATCAATGTGTAAAGCCTTGCATGTCATAATAACT 754  
 QY 300 GAAGAGACATGAATCAACCCATACAAAGTCATTTAAATGTACATTTGAAATTTCTCAAGA 359  
 Db |||||  
 755 TCAGAAACATGAAGAGACACATCTGGAGAGAAACCTTACAGTGTGATCAATGTATAATA 814  
 QY 360 AGCATTT---TATAACATCAATCTTTAAGACATCATATATATCTGTTCATGAAAAAAC 416  
 Db |||||  
 815 ASCCTTTGTATGAAGTATTTTCAAGTTTCATAAAAAACACATCTGGAGAGAAAC 874  
 QY 417 ATTAACGTGTAAACATGTAATAAGTTTTCACCTGCACCTTCAAAATTTAGCACACATAA 476  
 Db |||||  
 875 TTACAAATGTAATGAATGTGTAAAGCCTTTGCAGCACACATCTTCAAAAGTGCATAA 934  
 QY 477 ATTAAAAACATCATGTGTGATCTCTCTGTTATCAATGTGATCATCTGCTGTTGTTTAAAAA 536  
 Db |||||  
 935 AATAACACATCTGGAGAGAAAC---CTTACAAATGTAATCAATGTGTAAAGCCTTGC 991  
 QY 537 TTTCAAACTTGGTCAGTATTCATATTTCAATTTTCAATAAACAACATGCTCCAAACTTAATG 596  
 Db |||||  
 992 ATATCATAGTACACTCCAAGTACATCAAGAACACATACATCTGGAGAGAAACCTCATGATG 1051  
 QY 597 TCTTAAATCTGTAAAGTTGTGTGGGAAAAAGGTTTATCTTCCATATGTTAAAGTCA 656  
 Db |||||  
 1052 TGAGCAATGTGCAAGAGCCTTTGCAATCAAAAGTTATTTCCAAAGTACATAAAGATACA 1111  
 QY 657 TGATG 661  
 Db |||||  
 1112 TACTG 1116

RESULT 4  
 BC047646  
 LOCUS  
 DEFINITION Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone IMAGE:4828290), with apparent retained intron.

ACCESSION BC047646  
 VERSION BC047646.1 GI:29126804

KEYWORDS HTC.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2647)  
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalios D.E., Scherch A., Schein J.E., Jones S.J. and Marra M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE  
 JOURNAL MEDLINE  
 2238257  
 PUBMED 12477932

REFERENCE 2 (bases 1 to 2647)  
 Strausberg R.  
 Direct Submission  
 TITLE  
 JOURNAL Submitted (03-MAR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk

Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAK Plate: 90 Row: i Column: 7  
 This clone has the following problem: retained intron.

Location/Qualifiers  
 1. .2647  
 /organism="Homo sapiens"

/mol\_type="mRNA"  
 /db\_xref="taxon:9606"

/clone="IMAGE:4828290"  
 /tissue\_type="testis"

/clone\_lib="NIH MGC\_97"  
 /lab\_host="DH10B"

/note="Vector: pBluescript"

ORIGIN  
 Query Match 7.9%; Score 97.8; DB 11; Length 2647;  
 Best Local Similarity 50.1%; Pred. No. 1.7e-08;  
 Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;

QY 7 GAAAGTGAGAAACCAAAATCGATATCATCTTAAATATCTTCTTCTTCATCAGCTCC 66  
 Db |||||

1235 GAATGTGGCAAACTTTTAAACGATCTCAACTCTTACTACCCATAAGATAATTCATCT 1294  
 QY 67 AAAAGTATATTTGCATATGAAGGTGTGATAAAGCCTATAATCGACATCATTATTA 126  
 Db |||||

1295 GGAGAGAAACCTTACAATGTAAGATGTGTAAGCTTTTAAACGATCTTCAACCTT 1354  
 QY 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAATGTACAGTGCAGTGT 186  
 Db |||||

1355 ACTACCCATAGAAAAATTCATCTGAGAGAGAAACCTTTACAAATGT-----GAAGAATGT 1408  
 QY 187 GATAAGCATTTTTCAGAAAAATCACAATTTGGAAACACATATTTGATCACAATTCGAAAAA 246  
 Db |||||

1409 GGCAGAGCCTTTAAGCAGTCTCTCAACCTTACTACACATAGATATTTATCTGAGAG 1468  
 QY 247 AAACCATTTCCATTTGTTTCAGTGTGTGTAAGGGGTTTAAATTTCTCGACACACTTGAAGA 306  
 Db |||||

1469 AAACCTTACAAATGTAAAAATGTGAAAAAGCCTTTTAAACAGTCTGCACACCTTACCACA 1528  
 QY 307 CATGAATATCCCATACAAAGTCATTTAAATGTACATTTGAAATTTGCAAGAGCATTT 366  
 Db |||||

1529 CATGAGGTAATTCATCTGAGAGAAACCCCTACAAATGTGAAAAATGTGAAAAAGCCTTT 1588  
 QY 367 TATAA---ACATCAATCTTTAAGACATCATATATTTATCTGTTTCATGAAAAAACATTAA 423  
 Db |||||

1589 AATCATTTCTCACACCTTACTACATAGATAATTTCACTGAGAGAAACCTTACAAA 1648  
 QY 424 TGTAAACATGTAAATAAAGTTTTCATCTGACCTTCAAAATAGACACATTAATTAATA 483  
 Db |||||

1649 TGTAAAGATGTGTAAAGCTTTTAAACACTCTTCAACCTTACTTAAACATAGATAAT 1708  
 QY 484 CATCATGTGATCTCTCTGCTTATCAATGTGATCATCTGTTGTTTAAAAATTTCCAA 543  
 Db |||||

1709 CATCTGGAGAGAGCCTTTACAAATGTAAAGATGTGAAAAAGCCTTTTAAACCAATCTCA 1768

```

QY      544  ACTTGTCAGTATTACAAATTCATATATAAACAACACTGCATCCAAAACCTTAAATGTCTCTATAA 603
Db      1769  AAATCTTACTGAACATAAGAAAATTCATCTACGGAGAGAAACCCCTA---TGAATGTGAAGAAA 1825

QY      604  TGTGGTAAAGGTTGTCTGTGGAAAAAAGGTTTATCTTTCACATATGTTAAGTTCATGATGA 662
Db      1826  TGTGGCAAGGCTTTTAAACCAAGTCTCTCAATCTTACTAGACATAAGAAAAAGTTCATACAGA 1884

RESULT 5
AUI23448      745 bp  mRNA  linear  EST 01-AUG-2002
LOCUS      AUI23448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
DEFINITION
ACCESSION  AUI23448.1 GI:110948164
VERSION
KEYWORDS
SOURCE
ORGANISM  Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 745)
AUTHORS  Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Ishigai,T.
TITLE  HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
Suzuki,Y., Sugano,S., Ishigai,T.)
JOURNAL  Unpublished (2000)
COMMENT  Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5' - & 3' - end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM2000319"
/cell_type="textatocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RM2"
/notes="Vector: pME18SPL3; mRNA from uninduced NT2 neuronal
precursor cells"

ORIGIN
Query Match 7.9%; Score 97.6; DB 9; Length 745;
Best Local Similarity 50.2%; Pred. No. 2.1e-08;
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

QY      7  GAAAGTGACGAAACCAAAATCGATATCATCTTTTAATATATCTCTCTTTCATCATCGTCCC 66
Db      87  GAATGTGGCAAAACCTTTTAAACCGATCTCTCAACTCTTACTACCCATAAGATAATTCATACT 146

QY      67  AAAAGTATATTGGACATATGAGGGTGTGATTAAGCCCTATAATCGACATCATATTATA 126
Db      147  GGAGAGAAACCCCTACAAAATGTAAGAATGTGGTAAAGCTTTTAAACCGTTCTTCAACCCCT 206

QY      127  GAGCAACATTTAAGAACCCACAGTAATGATCGACCGGTATATAATGTACAGTGACAGTGT 186
Db      207  ACTTACCATAGAAAATTCATCTGAGAGAGAAACCTTACAAATGT-----GAAGAATGT 260

QY      187  GATAAGACATTTTCAGAAAATACATTTTGGAAACACATATTGTATCATCATTTCCGAAAAA 246
Db      261  GGCANAAGCCCTTAAAGCAGTCTCTCAACCTTACTTACACATAAGATAATTCATCTCGAGAG 320

```

```

QY      247  AAACCAATTCATTTGTCAGTGTGTGGTAAAGGGGTTAAATTTCTGCACAACACTTGAAGAAGA 306
Db      321  AAACCCCTACAAATGTAAAAAATGTGGAAAAAGCCCTTTTAAACCAAGTCTGCACACCTTTACTACA 380

QY      307  CATGAAATCACCCATACAAAGTCAATTTAAATGTACATTTGAAAAATTTGTCAAGAAGCAATTT 366
Db      381  CATGAGGTAAATTCATATCTGGAGAAACCCCTACAAATGTGAAAAATGTGGAAGAGCCTTT 440

QY      367  TATAA---ACATCAATCTTTAAGACATCATATATATCTGTTTCATGAAAAAACAATTAACG 423
Db      441  AATCATTTCTCACACCTTTACTACACATAAGATAATTTCTATCTGGAGAGAAACCTTTACAAA 500

QY      424  TGTAAACAGATGATAAAGTTTTCACTCGACCTTCAAAATTTAGCACACATATAAATTAATAA 483
Db      501  TGTAAAGAAATGTGTAAAGCTTTTAAACACTCTTCAACCCCTTACTAACAATAGATATTT 560

QY      484  CATCATGTTGGATCTCTGCTTATCAATGATCATCTCTGTTGTTTAAAAAATTTCCAA 543
Db      561  CATACTGGAGAGAAAGCCCTTACAAATGTAAAGAAATGTGAAAAAGCTTTTAAACCAATCTCTCA 620

QY      544  ACTTGGTCAGTATTACAATTTTCATATAAACAACACTGCATCCAAACTTAATATGTCCTAAA 603
Db      621  AAATCTTACTGAACATAAGAAAAATTCATCTGGAGAGAAACCCCTA---TGAATGTGAAGAAA 677

QY      604  TGTGTTAAAGGTTGTTGTGGAAAAAAGGTTTATCTTTCACATATGTTTAAGTCAT 657
Db      678  TGTGCGCAAGCTTTTAAACCAAGTCTCTCAAACTTACTAGACATAAGAAAAAGTCAT 731

RESULT 6
BC036394      2672 bp  mRNA  linear  HTC 19-NOV-2003
LOCUS      BC036394
DEFINITION  Homo sapiens zinc finger protein 85 (HFP4, HTP1), mRNA (cDNA clone
IMAGE:5259399), containing frame-shift errors.
ACCESSION  BC036394
VERSION  BC036394.1 GI:23025784
KEYWORDS  HTC.
SOURCE  Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2672)
AUTHORS  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaeetz,T.E., Brownstein,M.J., Udwin,T.B., Toehiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,F.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
PROC. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 2672)
Direct Submission
Strausberg,R.
Submitted (05-AUG-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
REMARK

```

COMMENT Contact: MGC help desk  
Email: cgapb-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
http://www.systemsbio.org  
contact: amadan@systemsbio.org  
Anup Madan, Jessica Fahay, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: iRAK Plate: 73 Row: a Column: 16  
This clone has the following problem: frame shifted.

FEATURES  
source  
1..2672  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5259399"  
/tissue\_type="Brain, hippocampus"  
/clone\_lib="NTH.MGC\_95"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

ORIGIN  
Query Match 7.5%; Score 92.4; DB 11; Length 2672;  
Best Local Similarity 54.1%; Pred. NO. 1.7e-07;  
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;  
QY 69 AAAGTATATTGCACATATGAAGGCTGCTAAGAGCTTAAAGCCTTAAATGACATCATTTATTAGA 129  
DB 1745 AGAGAAGCCTTACAAATGTAAGAAATGGAAGAGCTTTAAACCAATCCTCAAAACTTAC 1804  
QY 129 GCAACATTTAAGAACCCACAGATATGATCGACCGCTATAAATGTACAGTGGACGATTGTGA 188  
DB 1805 TGAACATAAGAAATTCATCTGGAGAGAAACCCCTATGATGT-----GAAAAATCTGG 1858  
QY 189 TAAGCATTTTTCAGAAAATCAGATTTGGAAACACATATGTCATCACTTCGCAAGAAAAA 248  
DB 1859 CAAAGCTTTTAAACAGTCTCTCAATCTTCTAGACATAGAAAGATGATACAGAGAGAA 1918  
QY 249 ACCATTCATTTTTCAGTGTGTGGTAAAGGGTTAAATCTCGACACACATTTGAAAGACA 308  
DB 1919 ACCTTACAATGTGAGAAATGTGGCAAGGTTTAAATGGCCCTCAACCTTACTATCCA 1978  
QY 309 TGAATACCCCATACAAAGTCATTTAAATGTACATTTGAAAATGTCAAGAGCATTTTA 368  
DB 1979 TAAGATAATTTCATCTGGAGAGAAACCATACAAATGTGAAGATGTGGCAAGCTTTTAA 2038  
QY 369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTCATGAAAAACATTAAAGTG 425  
DB 2039 CCAATCCTCAAACTTACCAACATAGAAATTCATCTGAGAGAAACCCCTACATG 2098  
QY 426 TAAACAATGTATAAAGTTTTCATCTGACCTCAAAATAGACACAAATAAATTAACA 485  
DB 2099 TGAAGATGTGGCAAGGCTTTAACAGTCTCTCAACCTTAACTAACAATAGAGATTCA 2158  
QY 486 TCATGGTGCATCTCT 501  
DB 2159 TACTGGAGAAAAACCT 2174

RESULT 7  
AQ005136  
LOCUS 591 bp DNA linear GSS 27-JUN-1998  
DEFINITION CIT-HSP-2290017.TF CIT-HSP Homo sapiens genomic clone 2290017,  
genomic survey sequence.  
ACCESSION AQ005136  
VERSION AQ005136.1 GI:3082581  
KEYWORDS GSS.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 591)  
AUTHORS Adams,M.D., Kounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
BUILDING Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Other GSSs: CIT-HSP-2290017.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21;  
Class: BAC ends.

FEATURES  
Location/Qualifiers  
1..591  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7150844"  
/db\_xref="taxon:9606"  
/clone="2290017"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

ORIGIN  
Query Match 7.4%; Score 92; DB 28; Length 591;  
Best Local Similarity 52.2%; Pred. No. 2.4e-07;  
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;  
QY 7 GAAAGTGCAAAACCAATCGATATCACTTTAAATATCTTCTTCTTCATCAGCTCC 66  
DB 69 GAATGGCAAGCCCTTCAACATTTTCACACCTTACACATAAGTAATTCATCT 128  
QY 67 AAAAAATATATTGACATATGAAGGCTGTGATAAGCCCTATAATCGACCATCATTTA 126  
DB 129 GGAGGAAGCCCTTCAATGTGAAGATGTGTAAAGCTTTTAAACCCCTTCAGCCCT 188  
QY 127 GAGCAACATTTAAGAACCCACAGTATGATGACCGTATAAATGTACAGTGGACGATGT 186  
DB 189 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCCTACAAATGT-----GAAGAATG 242  
QY 187 GATAAGCATTTTTCAGAAAAATCACATTTGGAACACACATATTGTATCACTTCGAAAA 246  
DB 243 GACAAAGCTTTTAAACGATTTCTCATACCTTACTAAACATAAGATAATTCATCTGGAG 302  
QY 247 AAACCATTCATTTGTCAGTGTGTGTAAGGGTTTAATCTCGACACACATTTGAAGA 306  
DB 303 AAATCTTACAAATGTGAACAATGTGGCAAGGCTTTAACTGGTCTTCAACCCCTTCAAAA 362  
QY 307 CATGAATCACCACATACAAAGTCATTTAAATGTACATTTGAAAAATTTGCAAGAAGCATTT 366  
DB 363 CATAAAGAATTCATCTGAGAGAAACCCCTACAAATGTGAAGATGTGGCAAGCCTTT 422  
QY 367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTTCATGAAAAACATTAAAG 423  
DB 423 AATGTCTCTCACACCTTACTACATAAGATGATTCATCTGAGAGAAACCCCTACAAA 482  
QY 424 TGTAAACAATGTAAATGAAGTTTTCATCTGACCTTCAAAATTAGCACCAACATTAATTA 483  
DB 483 TGTGAAGAATGTGGCAAGCCTTTTAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 542

```

QY      484 CATCATGGTG 493
Db      543 CATACTGGAG 552

RESULT 8
B99387 593 bp DNA linear GSS 26-JUN-1998
LOCUS   CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
DEFINITION genomic survey sequence.
ACCESSION B99387
VERSION   B99387.1 GI:3027197
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 593)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
BUILDING  (1998)
JOURNAL   Unpublished (1998)
COMMENT   Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES             Location/Qualifiers
     source
     1..593
         /organism="Homo sapiens"
         /mol_type="genomic DNA"
         /db_xref="GDB:7147044"
         /db_xref="taxon:9606"
         /clone="2281A10"
         /sex="Male"
         /cell_type="Sperm"
         /clone_lib="CIT-HSP"
         /note="Vector: pSelobAC11; Site_1: HindIII; Site_2:
HindIII"

ORIGIN
Query Match      7.3%; Score 91; DB 28; Length 593;
Best Local Similarity 54.4%; Pred. No. 3.7e-07;
Matches 231; Conservative 0; Mismatches 185; Indels 9; Gaps 2;

QY      69 AAAGTATATTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATATTATTAGA 128
Db      47 AAAGAAACCCCTACAAATGTGAGAGATGTGGCAAAGCTTTTAGCCAGTCCTCAACCCCTTAG 106

QY      129 GCACATTTTAAGAACCCACAGTAATGATCGACCGTATAATGTACAGTGGACGATTGTGA 188
Db      107 AAACATGATGATTAATTCATACCTCGGAGAAACCCCTCAAAATGT-----GAAGATGTGG 160

QY      189 TAAAGCATTTTTCAGAAAATTCACATTTGGAAAACACATATTGTATCACATTCGGAAGAAAAA 248
Db      161 TAAAGCTTTTAAAGTGTGCTCCACACCTTACTAGACATAAAGTAATTCATCTGAGAGAGAA 220

QY      249 ACCATTCCATTGTTAGTGTGGTAAAGGGTTAATTCGACACACATTCGAAAGACA 308
Db      221 ACCCTCAAAATGTGAAGATGTGGCAAGGCTTTTAAACCATTTTCTAGCCCTTAGGAACA 280

QY      309 TGAATCACCCTACCAAGTCAATTAAATGTACATTTGAAATTTGCAAGAGCATTTTA 368
Db      309 TGAATCACCCTACCAAGTCAATTAAATGTACATTTGAAATTTGCAAGAGCATTTTA 368

Db      281 TAAGATAAATTCATACTGGAAGAAACCCCTACAAATGTGAAGAATGTGGCAAGCTTTTAG 340
QY      369 TAAACATCAATCTTTT---AAGACATCATATATTATCTGTTCATGAAAAAACAATTAAAGCTG 425
Db      341 CCAGTCTCACTCTTAGAAAAACATGAGATAATTCACTAGGAGAGAAACCCCTACAAATG 400

QY      426 TAAACAATGTAAATGAAGTTTTCCTCGACCTTCAAATTTAGCACAACATATAATTAACA 485
Db      401 TGAAGAATGTGGTAAAGCTTTTAAAGTGTCTCTCAAAACCTTACTGTACATAAGGTAATCA 460
QY      486 TCATG 490
Db      461 TACTG 465

RESULT 9
BX414627 1163 bp mRNA linear EST 13-MAY-2003
LOCUS   BX414627 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP001VD12
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX414627
VERSION   BX414627.1 GI:30649903
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 1163)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqr@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9703.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSOCAP001DB06QP1&cluster=9703.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSOCAP001DB06QP1.
Location/Qualifiers
     1..1163
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="CSOCAP001VD12"
         /tissue_type="THYMUS"
         /clone_lib="Homo sapiens THYMUS"
         /note="Vector: PCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the PCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match      7.3%; Score 91; DB 13; Length 1163;
Best Local Similarity 50.1%; Pred. No. 3.5e-07;
Matches 289; Conservative 8; Mismatches 268; Indels 12; Gaps 3;

QY      89 AAGGGTGTGATAAAGCCTATAATCGACCATCATATTAGACACATTTAAGACCCACA 148
Db      121 AAGAATGTGGTAAAGCTTTTAAACCGATCTTCAACCCCTTACTAGCATAGAGRRTTCATA 180

QY      149 GTAATGATCGACCCGTATAAATGTACAGTGGACGATGTGATAAGCAATTTTTCAGAAAAAT 208
Db      181 CTGAGAGAAACCTTACAAATGT-----GAAGATGKGGCAAGCCCTTAAAGCAGTCT 234

QY      209 CACATTTGGAAACACATATTGTATCACTCCGAAAAAACCATTCCATTGTTCACTGT 268
Db      235 CAACCTTACTACATAAGATTAATTCATCTGGAGAGAAACCCCTACAAATGTAAGAAT 294

```





cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgcnhri.nih.gov](mailto:nisc_mgcnhri.nih.gov)  
 Akhter, N., Ayele, K., Beckstrom-Brenner, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Turgerson, C., Vogt, J.L., Walker, W.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 65 Row: 4 Column: 17  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 13386417  
 This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
 1..2253  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3674739"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor  
 arose spontaneously from a senescent normal mammary  
 (clone) outgrowth infected with the virus MMTV."  
 /clone\_lib="NCI CGAP\_Lu29"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 7.3%; Score 90.2; DB 11; Length 2253;  
 Best Local Similarity 51.9%; Pred. No. 4.6e-07;  
 Matches 256; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

QY 4 AGTGAAGTGCAGAAACCAATCATCTTTAATATCTCTTCTTCCTTCATCAGT 63  
 DB 823 AATCAATGTGATAAGGCTTTTCAACACACAGTCTCTTCAACATCATAGAGACACAT 882  
 QY 64 CCGAAAGTATATTTCACATATGAAGGTGTGATAAGCCCTTAATCGCACATCATTA 123  
 DB 883 ACCGGAAGGAAACCTTCAATGTATCAATGTGATAGAGCCCTTTCCCGTAAACATTTGT 942  
 QY 124 TTAGAGCAACATTTAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGCAGCAT 183  
 DB 943 GTGCABACCCTATAGGATACATCTGGAGAAACCCCTTCAATGTA-----ATCAA 996  
 QY 184 TGTGATAAGCATTTTCAGAAATCACAATTCATTGGAAACACATATTTGTATCATTTCCGAA 243  
 DB 997 TGTGATAAGCCCTTTTCAAGACAGTACTCTTCAAACTCATAGAGAACACATACCGGA 1056  
 QY 244 AAAAAACCATTCCTATGTTTTCAGTGTGTGTGTAAGGGGTTAATTCGCAACACTTGAAA 303  
 DB 1057 GAGAAACCTTCAATGTATCATGTTGATAAGCCCTTTCTGTTAAACATAGTCTCCAA 1116  
 QY 304 AGACATGAATACCCCTACAAAGTCAATTAATGTACATTTGAAATTTGCAAGACCA 363  
 DB 1117 ACACATAGGAATACATCTCGGGGAAACCCCTACAAATGTAATCAATGTGATAAGCC 1176  
 QY 364 TTTTATAACA---TCAATCTTTAAGACATCATATATTATCTGTTTCATGAAAAACATTA 420  
 DB 1177 TTTTCAACACATTTTCAACCTCAAAATCATATAAGAACACATCTAGAGAGATGCTTAC 1236  
 QY 421 ACGTGTAAACATGTATAAGTTTTCATCTCGACCTTCAAAATTTAGCACACATTAATTA 480  
 DB 1237 AAATGTAAATCAATGTGACAAAGCCCTTTTACGACATAGAAATCTTCAAGATTCATAGTGA 1296  
 QY 481 AAACATCATGGTG 493

DB 1297 AAACATCTGGAG 1309

## RESULT 13

BE161630  
 LOCUS  
 DEFINITION  
 MR3-HT0446-260300-201-f06 HT0446 Homo sapiens cDNA, mRNA sequence.  
 BE161630  
 VERSION  
 BE161630.1 GI:8624351  
 EST.  
 KEYWORDS  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 604)

## REFERENCE

## AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Mateukuma, A., Bada, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
 O'Hare, W.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20022663

10737800

PubMed

COMMENT

Lab: Simpson A.J.G.  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=8t2-MR3-HT0446-260>)  
 300-201-f06t3=2000-03-26t4=1

Seq primer: puc 18 forward

High quality sequence stop: 583.

## FEATURES

## source

1..604

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="HT0446"

/note="Organ: head neck; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No.196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

## ORIGIN

Query Match 7.2%; Score 89.2; DB 10; Length 604;  
 Best Local Similarity 53.7%; Pred. No. 8.1e-07;  
 Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;  
 QY 69 AAAGTATATTGCACATATGAAGGGTGTGATAAGCCCTATATCGACCATCATTTAGA 128  
 DB 102 AGAGAAACCTTCAAAATGTGAGAATGTGACAAAAGTTTTTAGTCGCTAATCACACCTTGA 161  
 QY 129 GCAACATTTTAAAGAACCCACAGTAATGATCGACCGGTATAAATGTACAGTGGACGATTGTGA 188  
 DB 162 AAGACATAGGAAATTCATCTCTGGAGAGAAACCCGACAAATGTAAAGT-----TTGTGA 215  
 QY 189 TAAAGCATTTTTCAGAAAATCATTTGGAAACACATATTTGTATCATCATTCGCAAAAAA 248  
 DB 216 CAAGGCTTTTTCAGACGCTGATTCACACCTGGCAACATATTTGTAATTCACCTGGAGAA 275

QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGTAAATTTCTCGACAACTTGAAAGACA 308  
 Db 276 ACCTTACAGTGTATAGTGTGGCAACACTTTGTTCAAAATTCATCTCTGTAATGCA 335  
 QY 309 TGAATATCCCATAC- --AAGTCAATTTAAATGTACATTTGAAATTTGCAAGAGCAAT 365  
 Db 336 TAAGTCAATTCATCTGAGAGAAATGTACAAGTGAATGAATGCGAAGTGTITTA 395  
 QY 366 TTATTAACATCATCTTTTAAGACATCATATATTATCTGTTCAATGAAAAAATTAACG 425  
 Db 396 TCACAAATCAACACTTGCATGTCATCATAGACTTCATCTGAGAGAAACCTTACAAG 455  
 QY 426 TAAACAATGTAATAAGTTTTCACGACCTTCAAAATTTAGCACAAATAAATAAACA 485  
 Db 456 TAATGAATGGCAAGGTTTTTAATGAATCAACCTTGAACATCATCATAGATTCA 515  
 QY 486 TCATGCTGATCTCT 501  
 Db 516 TATCGGAGAAAAACCT 531

RESULT 14  
 BC032590  
 LOCUS  
 DEFINITION Homo sapiens cDNA clone IMAGE:5502691, containing frame-shift errors.  
 ACCESSION BC032590  
 VERSION BC032590.1 GI:21619671  
 KEYWORDS HTC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 3609)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klatschner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Narusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schett, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Woley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932

2 (bases 1 to 3609)  
 Strausberg, R.  
 Direct Submission  
 Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 12477932  
 REFERENCE  
 2 (bases 1 to 3609)  
 Strausberg, R.  
 Direct Submission  
 Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)

Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Leric, P., Legaspi, R., Maduro, Q.D., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgel, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>  
 Series: TRAX Plate: 69 Row: 9 Column: 20  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13430873  
 This clone has the following problem: frame shifted.

## FEATURES

## source

1..3609  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5502691"  
 /tissue\_type="Eye, retinoblastoma"  
 /clone\_lib="NIH MGC\_67"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6"

## ORIGIN

Query Match 7.28; Score 89.2; DB 11; Length 3609;  
 Best Local Similarity 53.7%; Pred. No. 6.7e-07;  
 Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;  
 QY 69 AAGATATATTGACATATGAGGGTGTGATAAGGCTATAATCGACCATCATATTAGA 128  
 Db 587 AGAGAAACCTTACAAATGTGAGAATGTGACAAAGTTTGTAGTCGCTAATCACACCTGA 646  
 QY 129 GCACATTTAAGAACCCACAGTAATGATCGCGTATAATGTCACAGTGCAGCATTTG 188  
 Db 647 AAGCATAGAGAAATTCATCTGAGAGAAACCGTACAAATGTAAAGT-----TTGTGA 700  
 QY 189 TAAAGCATTTTTTCAGAAAAATTCACATTTGGAACACATATTGTATCATCATTCGGA 248  
 Db 701 CAAGGCTTTTCAGACGCTGATTCACACCTGGCACAAATATGTAATTCACACTGGAGAA 760  
 QY 249 ACCATTCATTTGTCAGTGTGTGTAAGGGTAAAGGGTTAATTCCTCGACACACTTGA 308  
 Db 761 ACCTTACAGTGAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTGTAATGCA 820  
 QY 309 TGAATCACCCTATAC---AAAGTCATTTAAATGTACATTTGAAAATTTGCAAGAAGCAT 365  
 Db 821 TAAGGTCATTCATCTGAGAGAGAAATGTTTACAAGTGAATGATGTGGCAAGGTTTTAA 880  
 QY 366 TTATTAACATCATCTTTAAGCATCATATATATCTGTTTCATGAAAAACATTAACGTG 425  
 Db 881 TCACAAATCAACCTTGCATGTCATAGATTCATCTGAGAGAAACCTTACAGTG 940  
 QY 426 TAAACAATGTAATAAGTTTTTCATCTGACCTTCAAAATTTAGCACAAATAAATAAACA 485  
 Db 941 TAATGAATGGCAAGGTTTTTAATTTGAAATCAAACTTGAACATCATCATAGATTCA 1000  
 QY 486 TCATGCTGGATCTCCT 501  
 Db 1001 TATCGGAGAAAAACCT 1016

## RESULT 15

AL705393  
 LOCUS  
 DEFINITION DXFZ5686M1835 r1 686 (synonym: hlccc3) Homo sapiens cDNA clone  
 DXFZ5686M1835 5', mRNA sequence.  
 ACCSSION AL705393  
 VERSION AL705393.1 GI:19688748

Search completed: May 9, 2004, 10:57:43  
Job time : 3623.38 secs

KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 535)  
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and  
Wismann,S.  
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and  
Wismann,S.)  
JOURNAL Unpublished (2001)  
COMMENT Contact: MIPS  
MIPS  
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wismann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wismann@dkfz-heidelberg.de;  
sequenced by MediGenomix (Martinsried/Germany) within the cDNA  
sequencing consortium of the German Genome Project. No sl sequence  
available.  
This clone (DKFZp686M1835) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.  
FEATURES  
Location/Qualifiers  
Source  
1..535  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZp686M1835"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="686 (synonym: hlcc3)"  
/note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB;  
cDNA-collection"

ORIGIN  
Query Match 7.2%; Score 88.6; DB 9; Length 535;  
Best Local Similarity 53.3%; Pred. No. 1.1e-06;  
Matches 237; Conservative 0; Mismatches 199; Indels 9; Gaps 2;  
QY 60 AGTCCCAAAAGTATTTGACATATGAGGGTGGATAAAGCCTATATGACCATC 119  
DB 86 ACATACTGGAAGAAATCTTTCAATGTAAAGATGTGAAAAGTCATTTGCATGCTTTC 145  
QY 120 ATTATTAGACCAATTTAAAGACCCACAGTAATGATCGACCGTATTAATGTACAGTGA 179  
DB 146 ACATTAGCTCAACATAAAGAAATTCATAGTGGAGAGAAACCTACAAATGTAAA---- 200  
QY 180 CGATTGTGATAAGCATTTTCAGAAAATCACAATTTGGAACACATATTGTATCACATC 239  
DB 201 -CAATGTGGAAAGCCTATATGAGACCTCAAAACCTTTCTACACATAAAGAATTCATAC 259  
QY 240 CGAAATAAAACCATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 299  
DB 260 TGAAGAAGAAACCCCTCAATGCGAAGAGTGTGGAAGAACCTTTAACCGGCTCTCACACCT 319  
QY 300 GAAAAGACATGAATCAACCCATACAAAGTCATTTAAATGTATCATTTGAAATTTGTCAAGA 359  
DB 320 TACTACACATAAGATAATTCATCTACTGGAAGAAACCCCTACAAATGTGAGGAGTGTGGCAA 379  
QY 360 AGCATTTTATAACAT---CAATCTTTAAGACATCATATATATCTGTTCAATGAAAAAC 416  
DB 380 AGCTTTTAAACCAATCTGCAACCTTACTACATAAGAGAAATTCATCTGGAGAGAAACC 439  
QY 417 ATTAAAGTGAACATGTAATAAAGTTTTCCTGACCTTCAAAATTTAGCACAACTAA 476  
DB 440 CTACAAATGTGAAGATGTGGAGAGCTTTTAGCCAGTCTTCAACCCCTTACTGCACATAA 499  
QY 477 ATTAAACATCATGTTGGATCTCCT 501  
DB 500 GATAATTCATCTGGAGAGAAACCT 524

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:14:51 ; Search time 58 Seconds  
(without alignments)  
2007.062 Million cell updates/sec

Title: US-09-831-804-3  
Perfect score: 2229  
Sequence: 1 MSEDSTKSISSLSSSSS.....PLVKARMOLLNPNETSVISR 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2229	100.0	412	3	AA993316
2	2225	99.8	412	3	AB73656
3	445	20.0	564	6	ABJ25655
4	445	20.0	564	6	ABJ26255
5	416.5	18.7	1212	4	ABG00399
6	403	18.1	1230	4	AAU30831
7	400	17.9	809	4	AAU30831
8	398	17.9	563	6	ABU96690
9	395.5	17.7	365	3	AA993317
10	395.5	17.7	409	4	AG75181
11	395.5	17.7	423	2	AA991305
12	395	17.7	622	4	AAU78947
13	395	17.7	631	4	AAU79931
14	395	17.7	632	4	ABG18386
15	395	17.7	719	4	ABG16953
16	393.5	17.7	1472	4	AAU31578
17	393	17.6	517	4	ABG03375
18	392.5	17.6	1050	4	ABG09685
19	392.5	17.6	1050	7	ADC32995
20	392	17.6	555	6	ADA54763
21	390.5	17.5	878	4	AU29528
22	390.5	17.5	1520	4	ABG27130
23	390	17.5	540	7	ABE65594
24	390	17.5	576	6	ADA54797
25	390	17.5	620	6	ABU96725

26	388.5	17.4	501	4	ABB71369	Abb71369 Drosophil
27	388	17.4	568	5	ABB79872	Abb79872 TRAF6-inh
28	387.5	17.4	872	7	ADC37535	Adc37535 Human nuc
29	386	17.3	727	4	ABG16954	Abg16954 Novel hum
30	385.5	17.3	474	6	ABU11782	Abu11782 Human MDD
31	385	17.3	577	5	AAE14680	AAe14680 Human tra
32	385	17.3	770	4	ABG10091	Abg10091 Novel hum
33	385	17.3	770	4	ABG14920	Abg14920 Novel hum
34	385	17.3	803	4	ABG20106	Abg20106 Novel hum
35	380	17.0	615	6	AAE37047	AAe37047 Human nuc
36	380	17.0	959	4	ABG01956	Abg01956 Novel hum
37	379	17.0	817	4	AAU40475	AAu40475 Human pol
38	378	17.0	613	4	AAU95862	AAu95862 Human pro
39	378	17.0	755	4	AAU40916	AAu40916 Human pol
40	378	17.0	907	4	AAU94428	AAu94428 Human pro
41	377	16.9	518	6	ABR41469	ABr41469 Human DIT
42	376	16.9	553	7	ADC31307	Adc31307 Human nov
43	376	16.9	582	6	ABR41407	ABr41407 Human DIT
44	376	16.9	675	5	ABP55423	ABp55423 Human zin
45	376	16.9	678	7	ADC31847	Adc31847 Human nov

## ALIGNMENTS

RESULT 1  
AA993316  
ID AA993316 standard; protein; 412 AA.  
XX AC AA993316;  
XX XX  
DT 04-SEP-2000 (first entry)  
XX XX  
DE A transcription factor designated CATFIIA.  
XX XX  
KW Transcription factor; CATFIIA; DNA-binding protein;  
KW ribosomal RNA 5S gene; fungal infection.  
XX OS Candida albicans.  
XX FH Key Location/Qualifiers  
FT Misc-difference 193  
FT Misc-difference 339 /note= "Ser encoded by CTG"  
FT Misc-difference 339 /note= "Ser encoded by CTG"  
XX XX  
FN WO200028037-A1.  
XX XX  
PD 18-MAY-2000.  
XX XX  
PF 09-NOV-1999; 99WO-FR002739.  
XX XX  
PR 10-NOV-1998; 98FR-00014147.  
XX (HMRI ) HOECHST MARION ROUSSEL.  
XX Bordon-Pallier F, Camier S, Sentenac A;  
XX WPI: 2000-376549/32.  
XX N-PSDB; AA15398.  
XX New nucleic acid encoding Candida albicans transcription factor, useful  
XX e.g. in screening for antimycotic agents and for immunization.  
XX Claim 12; Page 35-36; 45pp; French.

The present sequence represents a Candida albicans transcription factor, designated CATFIIA. The polypeptide is a DNA-binding protein, which is involved in initiating transcription of the ribosomal RNA 5S gene. The polynucleotide is used to screen for its specific inhibitors, potentially useful as antimycotic agents, to raise an antibody response that is protective against fungal infection and to raise antibodies. Such antibodies, as well as the polypeptides and polynucleotides are used in

CC compositions for diagnosing and treating fungal infections, e.g. by  
 CC detecting polymorphisms and mutations  
 XX  
 XX

SEQ Sequence 412 AA;

Query Match 100.0%; Score 2229; DB 3; Length 412;  
 Best Local Similarity 100.0%; Pred. No. 1e-167;  
 Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEDSTKSISSLSISSSSSPKKYICTYEGCDKAYNRPSLLQHLRTHSDRPPYKCTVD 60  
 DB 1 MSEDSTKSISSLSISSSSSPKKYICTYEGCDKAYNRPSLLQHLRTHSDRPPYKCTVD 60

QY 61 DCDKAFPRKSHLETHIVSHSEKPFHCGVCGKGVNSRQHLKRHEITHTKSFKCTPENCQE 120  
 DB 61 DCDKAFPRKSHLETHIVSHSEKPFHCGVCGKGVNSRQHLKRHEITHTKSFKCTPENCQE 120

QY 121 AFYKHQSLRHILSVHEKTLTCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNF 180  
 DB 121 AFYKHQSLRHILSVHEKTLTCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNF 180

QY 181 QTVSVLQPHIKQSHPKLCKPCGKGVCGKGLSHMLSHDDSTMIKIWTCDYCVGKFAK 240  
 DB 181 QTVSVLQPHIKQSHPKLCKPCGKGVCGKGLSHMLSHDDSTMIKIWTCDYCVGKFAK 240

QY 241 KNELVHYNIFHDGNIIPDDLKETEVRKLENNLDQGSKLNNLHELETKLVEEDEDEE 300  
 DB 241 KNELVHYNIFHDGNIIPDDLKETEVRKLENNLDQGSKLNNLHELETKLVEEDEDEE 300

QY 301 DSDLEKSDVRSDMSAQRSIKSFASLEGSKSVKLSINSKKKINCPKNCNDRMFSREY 360  
 DB 301 DSDLEKSDVRSDMSAQRSIKSFASLEGSKSVKLSINSKKKINCPKNCNDRMFSREY 360

QY 361 DLRRHLKWHDDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412  
 DB 361 DLRRHLKWHDDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412

RESULT 2

ABP73656

ID ABP73656 standard; protein; 412 AA.

XX AC ABP73656;

XX DT 30-JAN-2003 (first entry)

XX DE Candida albicans essential protein SEQ ID NO 7493.

XX KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;

XX KW signal transduction; DNA replication; cell division; growth;

XX KW proliferation; Candida albicans; fungicide; antifungal.

XX OS Candida albicans.

XX PN WO200253728-A2.

XX PD 11-JUL-2002.

XX PF 26-DEC-2001; 2001WO-US049486.

XX PR 29-DEC-2000; 2000US-0259128P.

XX PR 20-FEB-2001; 2001US-00792024.

XX PR 22-AUG-2001; 2001US-0314050P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;

XX DR WPI; 2002-566694/60.

XX DR N-PSDB; ABZ32206.

XX PT Constructing strains for identifying gene products as effective targets

XX PT for therapeutic intervention, by inactivating in the strain one allele of

PT a gene and placing other allele of the gene under conditional expression.  
 XX Claim 44; SEQ ID NO 7493; 167bp + Sequence Listing; English.  
 XX

The invention relates to constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying other allele by recombination of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M1) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthesis, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of C. albicans cells and for treating infection by C. albicans. The present sequence is that of an essential Candida albicans protein used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office

SEQ Sequence 412 AA;

Query Match 99.8%; Score 2225; DB 5; Length 412;

Best Local Similarity 99.8%; Pred. No. 2.1e-167;

Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEDSTKSISSLSISSSSSPKKYICTYEGCDKAYNRPSLLQHLRTHSDRPPYKCTVD 60

DB 1 MSEDSTKSISSLSISSSSSPKKYICTYEGCDKAYNRPSLLQHLRTHSDRPPYKCTVD 60

QY 61 DCDKAFPRKSHLETHIVSHSEKPFHCGVCGKGVNSRQHLKRHEITHTKSFKCTPENCQE 120

DB 61 DCDKAFPRKSHLETHIVSHSEKPFHCGVCGKGVNSRQHLKRHEITHTKSFKCTPENCQE 120

QY 121 AFYKHQSLRHILSVHEKTLTCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNF 180

DB 121 AFYKHQSLRHILSVHEKTLTCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNF 180

QY 181 QTVSVLQPHIKQSHPKLCKPCGKGVCGKGLSHMLSHDDSTMIKIWTCDYCVGKFAK 240

DB 181 QTVSVLQPHIKQSHPKLCKPCGKGVCGKGLSHMLSHDDSTMIKIWTCDYCVGKFAK 240

QY 241 KNELVHYNIFHDGNIIPDDLKETEVRKLENNLDQGSKLNNLHELETKLVEEDEDEE 300

DB 241 KNELVHYNIFHDGNIIPDDLKETEVRKLENNLDQGSKLNNLHELETKLVEEDEDEE 300

QY 301 DSDLEKSDVRSDMSAQRSIKSFASLEGSKSVKLSINSKKKINCPKNCNDRMFSREY 360

DB 301 DSDLEKSDVRSDMSAQRSIKSFASLEGSKSVKLSINSKKKINCPKNCNDRMFSREY 360

QY 361 DLRRHLKWHDDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412

DB 361 DLRRHLKWHDDNLQRIEFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412

RESULT 3

ABJ25655

ID ABJ25655 standard; protein; 564 AA.

XX AC ABJ25655;

XX DT 16-APR-2003 (first entry)

DE Aspergillus fumigatus essential gene protein #313.  
XX  
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
KW cancer; contamination; biofilm; antibody; immune response.  
XX  
XX Aspergillus fumigatus.  
OS  
XX  
XX W0200286090-A2.  
PN  
XX  
XX 31-OCT-2002.  
PD  
XX  
XX 23-APR-2002; 2002WO-US013142.  
PF  
XX  
XX 23-APR-2001; 2001US-0285697P.  
PR  
XX 27-APR-2001; 2001US-0287066P.  
PR  
XX 05-JUN-2001; 2001US-0295890P.  
PR  
XX 09-JUL-2001; 2001US-0303899P.  
PR  
XX 31-AUG-2001; 2001US-0316362P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
FA  
XX  
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
PI  
XX WPI; 2003-093124/08.  
XX  
XX New purified or isolated nucleic acids of essential genes of Aspergillus  
PT fumigatus, useful for treating or preventing infections by A. fumigatus,  
PT or for treating a non-infectious disease in a subject e.g. cancer.  
XX  
XX Disclosure; Page; 175pp; English.  
PS  
XX  
XX The invention relates to novel purified or isolated nucleic acids of  
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of  
CC the invention are used to treat or prevent infections by a pathogenic  
CC organism such as A. fumigatus, to treat a non-infectious disease in a  
CC subject (e.g. cancer), to prevent or contain contamination of an object  
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a  
CC biofilm comprising A. fumigatus. The polynucleotides are useful for  
CC expressing recombinant protein for characterization, screening or  
CC therapeutic use, as markers for host tissues in which the pathogenic  
CC organisms invade or reside, for comparing with the DNA sequence of A.  
CC fumigatus to identify duplicated genes or paralogues having the same or  
CC similar biochemical activity and/or function, for comparing with DNA  
CC sequences of other related or distant pathogenic organisms to identify  
CC potential orthologous essential or virulence genes, for selecting and  
CC making oligomers for attachment to a nucleic acid array for examination  
CC of expression patterns, for raising anti-protein antibodies, as an  
CC antigen to raise anti-DNA antibodies or to elicit another immune  
CC response, and for identifying polynucleotides encoding the other protein  
CC with which binding occurs or to identify inhibitors of the binding  
CC interaction. The polypeptides may be used to raise antibodies or to  
CC elicit immune response, as a reagent in assays designed to quantitatively  
CC determine levels of the protein in biological fluids, as a marker for  
CC host tissues in which pathogenic organism invade or reside, and to  
CC isolate correlative receptors or ligands in the case of virulence  
CC factors. This sequence represents a protein of one of the essential genes  
CC of Aspergillus fumigatus of the invention  
XX  
XX Sequence 564 AA:  
SO

	Query Match	20.0%;	Score 445;	DB 6;	Length 564;	
	Best Local Similarity	30.3%;	Pred. No. 1.8e-26;			
	Matches 125; Conservative	62;	Mismatches 164;	Indels	62;	Gaps 21;
QY	14 ISSSSSRP----	KKICTYECCDAKYRNP	ELLQHLETHNSRPYK	TCTVDDCDKAFFPKS	70	
	:: :	:: :: :	:: :: :	:: :: :	:: :: :	
Db	71 ISTTSAKYSELTKHRC	FPDGGCTKAFNRPARLO	EHLRSNNERIPKCTPEEC	DKTFLAS	130	
	:: :	:: :: :	:: :: :	:: :: :	:: :: :	
QY	71 HLETHIVS-HSEKKPFCS-	-VCCKGVNSQHLKRHEITH-	-TKSPKCT-PENCCEAYFK	124		
	:: :	:: :: :	:: :: :	:: :: :	:: :: :	
Db	131 HLNHIIKSAHTGVRDY	VCDRPGCKSVFGSRURRLAAHDGR	KDYKTCETPPCNETFRK	190		
	:: :	:: :: :	:: :: :	:: :: :	:: :: :	
QY	125 HQSLRHHTLSVH--EKTILT	C-----KOCNVFTFRPSKLAKHLKHGHGS	PAYQCCH-	173		
	:: :	:: :: :	:: :: :	:: :: :	:: :: :	

Db	191	HSTLQKHINTAHLKQKPFQCPHTDPTSGQCKTMAFDAGHLRAHESRIH-TEKRFSCTEC	2419
Qy	174	----PGCFRNFTWSVLOPHIKQGHPLKCPKCGKGVGGKGLSSHM-LSHDDSTM--IK	2246
Db	250	SQAEGAEATFTYALLQAHIRSVP--QCPNCALTCATSRELRRHLEVAHGDVSLERK	3008
Qy	227	IWTCDY--CDVGKFAKKNELVEHYNIEHDGNI-----PDLLKETEYKKLENLLDQ----	2775
Db	309	IFPCTVPGGD--RSFTYKGNLTVHRTVHQEKRFVCGETDLSSKKVSGWNN--DNGCGK	3655
Qy	276	--GSKNNLHELETEKLVBEDEDEBSL-----DEKRSVDRSDSMSAQRISKFTASLE	3299
Db	366	RYGSKALAEHRTTAHLGYONAKAERRQRLGITRDROHSTATSPGWSA-----LAALT	418
Qy	330	GSKSVSKLISNGKKINCPKNCDRMFSREYDLRHL--KWHDDNLQRIEBSFL	380
Db	419	GEGYAE-----TGRHIACLVESCPRHTRHYDLVWVMGKHHSFSEETRLFL	467
RESULT 4			
ID	ABU26255	standard; protein; 564 AA.	
AC	ABU26255;		
CC	XX		
DD	XX		
EE	XX	16-APR-2003 (first entry)	
FF	XX	Aspergillus fumigatus essential gene protein #913.	
GG	XX	Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;	
HH	XX	cancer; contamination; biofilm; antibody; immune response.	
II	XX	Aspergillus fumigatus.	
JJ	XX	WO200286090-A2.	

## RESULT 4

ABJ26255	
ID	ABJ26255 standard; protein; 564 AA.
XX	
AC	ABJ26255;
XX	
DT	16-APR-2003 (first entry)
XX	
DE	Aspergillus fumigatus essential gene protein #913.
XX	
KW	Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW	cancer; contamination; biofilm; antibody; immune response.
XX	
OS	Aspergillus fumigatus.
XX	
PN	WO200286090-A2.
PN	
PD	31-OCT-2002.
XX	
XX	
XX	23-APR-2002; 2002WO-US013142.
XX	
XX	23-APR-2001; 2001US-0285697P.
PR	27-APR-2001; 2001US-0287066P.
PR	05-JUN-2001; 2001US-0295890P.
PR	09-JUL-2001; 2001US-0303899P.
PR	31-AUG-2001; 2001US-0318362P.
XX	
XX	(ELIT-) ELITRA PHARM INC.
PA	
XX	
PI	Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX	
XX	WPI; 2003-093124/08.
XX	
XX	
PT	New purified or isolated nucleic acids of essential genes of Aspergillus
PT	fumigatus, useful for treating or preventing infections by A. fumigatus;
PT	or for treating a non-infectious disease in a subject e.g. cancer.
XX	
XX	Disclosure; Page: 175pp; English.

Disclosure; Page; 175pp; English.

The invention relates to novel purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterization, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential virulence genes, for selecting and

CC making oligomers for attachment to a nucleic acid array for examination  
 CC of expression patterns, for raising anti-protein antibodies, as an  
 CC antigen to raise anti-DNA antibodies or to elicit another immune  
 CC response, and for identifying polynucleotides encoding the other protein  
 CC with which binding occurs or to identify inhibitors of the binding  
 CC interaction. The polypeptides may be used to raise antibodies or to  
 CC elicit immune response, as a reagent in assays designed to quantitatively  
 CC determine levels of the protein in biological fluids, as a marker for  
 CC host tissues in which pathogenic organism invade or reside, and to  
 CC isolate correlative receptors or ligands in the case of virulence  
 CC factors. This sequence represents a protein of one of the essential genes  
 CC of Aspergillus fumigatus of the invention

XX Sequence 564 AA;

Query Match 20.0%; Score 445; DB 6; Length 564;  
 Best Local Similarity 30.3%; Pred. No. 1.8e-26;  
 Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

QY 14 ISSSSSRP---KKYICTYEGCDKAYNRPSLLEOHLRTHSDNRPYKCTVDDCDKAFPKS 70  
 Db 71 ISTTSKYPSELKTHRCPPDCCTKAFNPAQLQHLASHNNERIFKCTFECDKTLRAS 130  
 QY 71 HLETHIVS-HSEKPFHGS--VCGKGVNSROHLKRHEITH--TKSFKCT-FENCQEAFLK 124  
 Db 131 HLNHHIKSAHTGVADYVCDRPGCGKSFVTGSLRHLAAHDGRDKYRCTEYPPCNETFRK 190  
 QY 125 HQSLRHILSVH--EKLTC-----KOCNKVFTRPKSLAQHKLKHGGSPAYOCDS- 173  
 Db 191 HSTLQKHIMTAHLKQKPPQCHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFCTEC 249  
 QY 174 ----PGCFNFQVSWVLOFHIKQSHPLKPKCKGKGVGKKGSSSM--LSHDDSTM--IK 226  
 Db 250 SOHAEAGAEATPFTYALLQAHIRSVHPP-QCFNCALTCATSRLELRHLEVAHGDVSLERK 308  
 QY 227 IWTCDY--CDVGKFAKKNELVEHYNIPHDGNI-----PDLKETEKKLENLLDQ--- 275  
 Db 309 IFPCTVPCD-RSTKKNLTVHRTVHQEKFPVCGETDLSKKVSGWNN--DNGCGK 365  
 QY 276 --GSKLNNLHELETKLVEDEDEEDSL-----DEKRSVRSDSMSAORSKSFASLE 329  
 Db 366 RYGSKLALEEHRTAHLGQYNAKAEQRQLGITRDRQHSFATSPGVSA-----LAALT 418  
 QY 330 GSKSVKLSNGKKNCPXNCDRMFSRSDYDLRRLH--KWHDDNLQRTESPL 380  
 Db 419 GEGYAE-----TGRHACLVECPHRRDYDLVWMSGKHSESEETRDLEL 467

RESULT 5  
 ABG00399  
 ID ABG00399 standard; protein; 1212 AA.

XX ABG00399;

AC ABG00399;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #390.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;  
 PI WPI, 2001-639362/73.  
 XX N-FSDB; AAS64586.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 30758; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1212 AA;

Query Match 18.7%; Score 416.5; DB 4; Length 1212;  
 Best Local Similarity 32.8%; Pred. No. 9.1e-24;  
 Matches 116; Conservative 45; Mismatches 148; Indels 47; Gaps 15;

QY 20 SRPKYICTYEGCDKAYNRPSLLEOHLRTHSDNRPYKCTVDDCDKAFPKSHLETHIVSH 79  
 Db 861 TRETKYKC--EECGKAFSQPSHLTTHKRMHTGKPKYC--EECGKAFSQSSLTTHKIIH 916  
 QY 80 SEKKPHFCSVCGKGVNSROHLKRHEITH--KSFKCTFENCQEAFLKQSL-RHILSVH 136  
 Db 917 TGEKPKYCEGKAFKRSSTLTTEHKIHTGKPKYC--EECGKAFSQSSLTTHRTRMTG 974  
 QY 137 EKLTCCKQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKNFQVSWVLOFHQSHPK 196  
 Db 975 EKPCKEECGKAFNRSSKLTTHKIHTGKPKYKCEE--CGKAFISSSTLNGH-KRIHTR 1030  
 QY 197 ---LKCPKCGKGVGKGLSSHMLSHDDSTMKIWTCDYCVGKFAKKNELVEHYNIEHD 253  
 Db 1031 EKPCKEECGKAFSQSSLTTRHKLRLH---TGEKPKYKCEE--GKAFKSSALTTHKIIHT 1085  
 QY 254 GNIPDOLLKETEYKLENLLDQSKLNNLHLETEKLVKVEDEDEEDSLDEKRSVRS 313  
 Db 1086 GE-----KPYKCEKCGKAFNQSSILTNHKKIHTTTPKIH-----TREK 1123  
 QY 314 SMSAQRISKSFASLEGSVSKLSNGKKNKINCPCNMDRMFSREYDLRHLKWH 369  
 Db 1124 PYPKCEGKFSNRSSTFTK--HKVIHTGVKLYKC--EECGKSFWSALTTHKIIH 1175

RESULT 6

AAU30831

ID AAU30831 standard; protein; 1230 AA.

XX AAU30831;

AC AAU30831;

DT 18-DEC-2001 (first entry)

```

XX DE Novel human secreted protein #1322.
XX KW Human; vaccination; gene therapy; nutritional supplement;
XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX PP 16-APR-2001; 2001WO-US008656.
XX PR 18-APR-2000; 2000US-00552929.
XX PR 26-JAN-2001; 2001US-00770160.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX PI WPI; 2001-611725/70.
XX DR Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy.
XX PS Claim 20; Page 355; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising the
XX CC nucleic acids encoding the polypeptides and cells genetically engineered
XX CC to express them are also useful for producing the proteins. The proteins
XX CC are useful in genetic vaccination, testing and therapy, and can be used
XX CC as nutritional supplements. They may be used to increase stem cell
XX CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
XX CC and/or nerve tissue growth or regeneration; immune suppression and/or
XX CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
XX CC AAU29510-AAU3304 represent the amino acid sequences of novel human
XX CC secreted proteins of the invention
XX SQ Sequence 1230 AA;

Query Match 18.1%; Score 403; DB 4; Length 1230;
Best Local Similarity 32.9%; Pred. No. 1.1e-22;
Matches 118; Conservative 45; Mismatches 144; Indels 52; Gaps 17;

QY 20 SRPKYICTYEGCDKAYNRPSLLPQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVSH 79
DB 891 TREAPYKC--EECGKAFSQSHLTTHKRMHTGKRPYKC--EECGKAFSQSSTLTTHKIIH 946
QY 80 SEKXPFFHCVCCKGNGRQHLKRHEIHT--KSPKCTFENCQEAIFYKHQSL-RHHILSVH 136
DB 947 TGEKPYKCECCGKAFKRSSTLTETHEKIHTGKPYKC--EECGKAFSQSSTLTTHKRMHTG 1004
QY 137 EKTLTCKCNKVFTRPSKLAQHLKHGGSPAYOCDPHGCFKFNQFQVSWLVQFHKQSHPK 196
DB 1005 EKPVCCECGKAFKRSSTLTTHKIIHTGKEP-YKCEE--CGKAFISSTLNGH-KRIHTR 1060
QY 197 LKCPKCGKGVGK-----KGLSSHMLSHDDSTMIKIWTCDYCDVGFKAKNELVEHNYI 250
DB 1061 EKPYKC-EGC-GKAFSQSFNTLTGKRLH---TGEKPYKQCEC--GKAFKSSALTTHKI 1113
QY 251 PHDGNIPDDLKETEVEVKLENLLDOGSKLNNLHLETEKLVKEEDEDDEEDSLDEKESDV 310
DB 1114 IHTGCE-----KPYKCECCGKAFNQSSILTTHKIIHTTPKIH-----T 1151

```

```

QY 311 RSDSMSAQRISIKFTASLEGSKSVKLSNCKKINCPKNCNCDRMESREYDLRRHLKWH 369
DB 1152 REKPYKKECGKSFNRSSTFTK--HKVIHTGVKLYKC--EECGKSFVWSSALTTHKIIH 1206

RESULT 7
ID AAM38689 standard; protein; 809 AA.
XX AC AAM38689;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 1834.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US034263.
XX PR 23-DEC-1999; 99US-00471275.
XX PR 21-JAN-2000; 2000US-00488725.
XX PR 25-APR-2000; 2000US-00552317.
XX PR 20-JUN-2000; 2000US-00598042.
XX PR 19-JUL-2000; 2000US-00620312.
XX PR 03-AUG-2000; 2000US-00653450.
XX PR 14-SEP-2000; 2000US-00662131.
XX PR 19-OCT-2000; 2000US-00693036.
XX PR 29-NOV-2000; 2000US-00727344.
XX PA (HYSE-) HYSEQ INC.
XX KW Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX KW Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX KW Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX N-PSDB; AAI57845.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX Example 3; SEQ ID NO 1834; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX Sequence 809 AA;

Query Match 17.9%; Score 400; DB 4; Length 809;

```

The invention describes a novel human isolated nucleic acid-associated polypeptide (NAAP). The polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of NAAP, such as cell proliferative (e.g. cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS, allergies), and developmental (e.g. Hypothyroidism, Cushing's syndrome) disorders, or infections. These are also useful in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of NAAP. The NAAP or its fragments are useful in screening compounds for effectiveness as agonist or antagonist of the polypeptides, or in altering the expression of the target polynucleotide and compounds that specifically bind to or modulate the activity of the polypeptide. The microarray is useful in monitoring or measuring protein-protein interactions, drug-target interactions, and gene expression profiles. This is the amino acid sequence of a novel human nucleic acid-associated protein (NAAP).

157

AC AAY93317;  
 XX 04-SEP-2000 (first entry)  
 XX A human transcription factor designated htIIIA.  
 XX Human; transcription factor; htIIIA; DNA-binding protein; transcription;  
 XX ribosomal RNA 5S gene; transcriptional control; cancer.  
 XX Homo sapiens.  
 XX WO200028024-A1.  
 XX 18-MAY-2000.  
 XX 09-NOV-1999; 99WO-FR002738.  
 XX 10-NOV-1998; 98FR-00014146.  
 XX (HMRI ) HOECHST MARION ROUSSEL.  
 XX Bordon-Pallier F, Rocher C;  
 XX WPI; 2000-387419/33.  
 XX N-ESDB; AAA15405.  
 XX New nucleic acid encoding human transcription factor IIA, useful for  
 XX treatment and diagnosis of cancer and inherited disease.  
 XX Claim 10; Page 40-41; 49pp; French.  
 XX The present sequence represents a human transcription factor designated  
 XX htIIIA. The polypeptide is probably a DNA-binding protein probably  
 XX involved in initiating transcription of the gene for ribosomal RNA 5S and  
 XX maintaining the stability of transcription of other control genes. The  
 XX htIIIA polynucleotides and polypeptides are used to make therapeutic or  
 XX diagnostic compositions for diseases associated with disorders of  
 XX transcriptional control, particularly cancer or other inherited diseases.  
 XX The htIIIA polynucleotide can also be used to detect anomalies in gene  
 XX transcription, particularly for diagnosis of inherited disease, also for  
 XX studying diseases involving htIIIA  
 XX Sequence 365 AA;  
 Query Match 17.7%; Score 395.5; DB 3; Length 365;  
 Best Local Similarity 31.1%; Pred. No. 8.2e-23;  
 Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;  
 Qy 8 KSISL-----ISSSSSR-----PKYICTYEGCDKAYNRPSSLLEOHLRTHSND 53  
 Db 9 ESVSLLTIADAFIAGESSAFTPPRALPRFICSPDCSANYSKAWKLDALCKHTGER 69  
 Qy 54 PYKCTVDDCDKAFPRKSHLETHIVSHSEKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107  
 Db 69 PFVCDYEGCGKAFIRDYHLSRHLTHGTGKPFVCAATGCDQKFNKSNLKKHFKHENO 128  
 Qy 108 TKSEKCTFENCQEAIFYKHSLR-HHILSVHEKTLTKQ--CNKVTRPSKLAQHLKHKG 164  
 Db 129 QKQYICSFEDCKKTFKXHQQLKHCQHTNEPLFKCTQEGCGKHFPASPSKLKHAHAHEG 188  
 Qy 165 GSPAYQCDHPGCFKFNFTWSVLQPHIQSHPK----- 196  
 Db 189 ----YVC-QKGCSPVAKTWTELLGHVRETHKEBTLCEVCRKTPKRDYLLQHKMTHAPER 243  
 Qy 197 --LKCPK--CGKGVGKKGSSHMLS-HDDSTMIKIWTCDYDVGK-FAKGNELVHYNI 250  
 Db 244 DVCRCREGCGRTVTVFNLSHLSFHEBS---RPFVCEHAGCGKTFAMKQSLTRH-AV 299  
 Qy 251 FHDGNIPIDDLKETEYVK 268  
 Db 300 VHD---PDKKQMKLVKK 314

RESULT 10  
 AAG75181  
 XX AAG75181 standard; protein; 409 AA.  
 XX AAG75181;  
 XX 03-SEP-2001 (first entry)  
 XX Human colon cancer antigen protein SEQ ID NO:5945.  
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 XX colorectal carcinoma; chromosome 13.  
 XX Homo sapiens.  
 XX WO200122920-A2.  
 XX 05-APR-2001.  
 XX 28-SEP-2000; 2000WO-US026524.  
 XX 29-SEP-1999; 99US-0157137P.  
 XX 03-NOV-1999; 99US-0163280P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
 XX WPI; 2001-235357/24.  
 XX N-PSDB; AAH34586.  
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 XX useful for preventing, diagnosing and/or treating colorectal cancers.  
 XX Claim 11; Page 7452-7454; 9803pp; English.  
 XX AAH32943 to AAH37195 and AAG73514 to AAG77798 represent human colon  
 XX cancer-associated nucleic acid molecules (N) and proteins (P), where the  
 XX proteins are collectively known as colon cancer antigens. The colon  
 XX cancer antigens have cytostatic activity and can be used in gene therapy  
 XX and vaccine production. N and P may be used in the prevention, diagnosis  
 XX and treatment of diseases associated with inappropriate P expression. For  
 XX example, N and P may be used to treat disorders associated with decreased  
 XX expression by rectifying mutations or deletions in a patient's genome  
 XX that affect the activity of P by expressing inactive proteins or to  
 XX supplement the patient's own production of P. Additionally, N may be used  
 XX to produce the colon cancer-associated Ps, by inserting the nucleic acids  
 XX into a host cell and culturing the cell to express the proteins. N and P  
 XX can be used in the prevention, diagnosis and treatment of colorectal  
 XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent  
 XX sequences used in the exemplification of the present invention. N.B.  
 XX Pages 666 to 682 and page 7053 of the sequence listing were missing at  
 XX time of publication, meaning no sequences are present for SEQ ID NO:1027  
 XX to 1052, 7921 and 7922  
 XX Sequence 409 AA;  
 Query Match 17.7%; Score 395.5; DB 4; Length 409;  
 Best Local Similarity 31.1%; Pred. No. 9.6e-23;  
 Matches 99; Conservative 54; Mismatches 95; Indels 69; Gaps 16;  
 Qy 8 KSISL-----ISSSSSR-----PKYICTYEGCDKAYNRPSSLLEOHLRTHSND 53  
 Db 53 ESVSLLTIADAFIAGESSAFTPPRALPRFICSPDCSANYSKAWKLDALCKHTGER 112  
 Qy 54 PYKCTVDDCDKAFPRKSHLETHIVSHSEKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107  
 Db 113 PFVCDYEGCGKAFIRDYHLSRHLTHGTGKPFVCAANGCDQKFNKSNLKKHFKHENO 172  
 Qy 108 TKSEKCTFENCQEAIFYKHSLR-HHILSVHEKTLTKQ--CNKVTRPSKLAQHLKHKG 164  
 Db 173 QKQYICSFEDCKKTFKXHQQLKHCQHTNEPLFKCTQEGCGKHFPASPSKLKHAHAHEG 232

QY 165 GSPAYQCDHPGCFKQNFQTSVLQFHFKOSHHPK----- 196  
Db 233 ----YVC-QKGCSPVAKTWTLLKHVRETHKEEILCEVCRKTFKFKDYLLKQHMKTAPER 287  
QY 197 --LKCPK--CGKGCVGKGLSHMLSHDDSTMKIWTCDYDVGK-FAKQNELVEHNYI 250  
Db 288 DVCRCFREGCGRTYTTVFNLSHLSFHES---RPFVCEHAGCGKTFAMKQSLTRH-AV 343  
QY 251 FHDGNIPDDLKETEYVK 268  
Db 344 VHD---PDKKKMKLVKK 358  
RESULT 11  
ID AAR91305 standard; protein; 423 AA.  
XX AAR91305;  
XX 07-JUL-1996 (first entry)  
XX Transcription factor-IIIa.  
XX Human; transcription factor-IIIa; hTFIIIA; DNA binding protein; ribosome;  
KW zinc finger; diagnostic; probe; transcription control; antitumour;  
KW cancer; therapy.  
XX Homo sapiens.  
XX EP704526-A1.  
XX 03-APR-1996.  
XX 05-SEP-1995; 95EP-00113908.  
XX 05-SEP-1994; 94JP-00211022.  
XX (SAXA ) OTSUKA PHARM CO LTD.  
XX Fujiwara T, Takeda S, Shimada Y, Ozaki K, Shin S;  
XX WPI; 1996-173033/18.  
XX N-PSDB; AAT14037, AAT14038.  
XX Human Transcription Factor III A gene - useful in regulation of  
PT transcription and for diagnosis and treatment of e.g. cancer related  
PT diseases.  
XX Claim 1; Page 9-10; 17pp; English.  
XX The sequence represents human transcription factor-IIIa (hTFIIIA), a DNA  
CC binding protein which is necessary for the initiation of 5S RNA gene  
CC transcription, binding to an internal control region of the 5S gene. The  
CC protein contains 9 zinc finger domains, which are homologous to the C2H2  
CC finger domains of Xenopus TFIIIA, except for the 6th finger domain, which  
CC has only 3 amino acid residues between 2 cysteine residues, instead of 5  
CC amino acid residues for Xenopus TFIIIA. The protein (optionally in  
CC recombinant form) and encoding gene may be used in diagnosis,  
CC identification or therapy of hereditary diseases such as cancer, or other  
CC diseases resulting from abnormal transcriptional control, and to analyse  
CC the mechanisms involved in their activity  
XX SQ Sequence 423 AA;  
Query Match 17.7%; Score 395.5; DB 2; Length 423;  
Best Local Similarity 30.8%; Pred. NO. 1e-22;  
Matches 99; Conservative 56; Mismatches 95; Indels 69; Gaps 16;  
QY 8 KSISSL-----ISSSSSR-----PKYICTYEGCDKAYNPSSLLEQLRTHSND 53  
Db 67 ESVSLTIADTAAGESSAPTFRPALPRFRFCSPDCGSANYSKAWKLDLHLCKHTGER 126  
QY 54 PYCKTVDCCDKAPFRKSHLETHIVSHSEKKPFCVS--CGKGVNQRQLKRH-EITH--- 107

Db 127 PFVCDYEGGKAFIRDYHLRHLTHGTGKPFVCAAGCDQKENTKSNLKKGFERKHENQ 186  
QY 108 TKSFKCTFFENCQAFYKHQSLR-HHLSVHEKTLTCKQ--CNKVFTRPSPKLAHQHKLKHG 164  
Db 187 QKQYISFSDCKTFFKHQOMKIHQONTNEPLFKCTQEGCGKHFPASPKLKEHAKAHG 246  
QY 165 GSPAYQCDHPGCFKQNFQTSVLQFHFKOSHHPK----- 196  
Db 247 ----YVC-QKGCSPVAKTWTLLKHVRETHKEEILCEVCRKTFKFKDYLLKQHMKTAPER 301  
QY 197 --LKCPK--CGKGCVGKGLSHMLSHDDSTMKIWTCDYDVGK-FAKQNELVEHNYI 250  
Db 302 DVCRCFREGCGRTYTTVFNLSHLSFHES---RPFVCEHAGCGKTFAMKQSLTRH-AV 357  
QY 251 FHDGNIPDDLKETEYVK 268  
Db 358 VHD---PDKKKMKLVKK 372  
RESULT 12  
ID AAM78947 standard; protein; 622 AA.  
XX AAM78947;  
XX 06-NOV-2001 (first entry)  
XX Human protein SEQ ID NO 1609.  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX Homo sapiens.  
XX WO200157190-A2.  
XX 09-AUG-2001.  
XX 05-FEB-2001; 2001WO-US004098.  
XX 03-FEB-2000; 2000US-00496914.  
XX 27-APR-2000; 2000US-00560875.  
XX 20-JUN-2000; 2000US-00598075.  
XX 19-JUL-2000; 2000US-00620325.  
XX 01-SEP-2000; 2000US-00654936.  
XX 15-SEP-2000; 2000US-00663561.  
XX 20-OCT-2000; 2000US-00693325.  
XX 30-NOV-2000; 2000US-00728422.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI; 2001-476283/51.  
XX N-PSDB; AAK52080.  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX Claim 20; Page 3945-3946; 6221pp; English.  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
SQ Sequence 622 AA;

Query Match 17.7%; Score 395; DB 4; Length 622;  
Best Local Similarity 29.7%; Pred. No. 1.9e-22;  
Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;  
QY 23 KKYICTVEGCDKAYNRPSLLQHLTHNSDRPYKCTVDDCDKAPFRKSHLETHIVSHSEK 82  
DB 201 KPYIC--ECCGKAFKYSALNTHKRIHTGKPYKC--DKCDKAFIASSTLSKHEIHTGK 256  
QY 83 KPFFCSVCGVGNRSQHLKRHEIHT--KSFKCTFENCQAFYKHQSLRHH-----131  
DB 257 KPYKCECGKAFNQSSTLTGKHIHTGKPYKC--ECCGKAFNQSSTLTGKHIHTGK 314  
QY 132 -----ILSVH-----EKTLTCKQCNKVTRPSKLAQHLKH-----162  
DB 315 YVCECGKAFKYSRILTHKRIHTGKPYKCNKCGKAFIASSTLSRHEFTGKHKHYKCE 374  
QY 163 -----HGGSPAYQCDHPGCFKFNQFQWVLOPHIKOSH---PKLKCPK 201  
DB 375 ECGKAFIWSVLTGKHIHTGKPYKCEE--CGKAFKYSSTLSH--KRSHTGKPYKCE 431  
QY 202 CGKGCVGKGLSSHMLSHDDSTMIKIWTCDYGVGK--FAKKNELVEHYNIHFDGNIPDDL 260  
DB 432 CGKAFVASTLSKHEIHT--TGKKPYKCEC--GKAFNQSSTLTGKHIHTGK 482  
QY 261 LKETEYKLENLLDQGSKLNNLHELET--EKLKVEDEDEDEDS---LDEKRSVRSDSM 315  
DB 483 YKCECGKAFN---QSSSLTGKHIHTGKPYKCECGKAFNQSSTLTGKHIHTREKPY 539  
QY 316 SAORSIKSFASLEGSKSVKLSNSGKKNPCPNNDMFRSREYDLRRLKWH 369  
DB 540 KCECGKAF--HLSTHLTTHKILHTGKPYRC--RECCKAFNHSATLSHKKI 589

RESULT 13  
AAM79931  
ID AAM79931 standard; protein; 631 AA.  
XX  
AC AAM79931;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 3577.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US004098.  
XX  
PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654936.  
PR 15-SEP-2000; 2000US-00653561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.  
PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI; 2001-476283/51.  
DR N-PSDB; AAK53064.  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX  
PS Claim 20; Page 388; 6221pp; English.  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX

SQ Sequence 631 AA;  
Query Match 17.7%; Score 395; DB 4; Length 631;  
Best Local Similarity 29.7%; Pred. No. 1.9e-22;  
Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;  
QY 23 KKYICTVEGCDKAYNRPSLLQHLTHNSDRPYKCTVDDCDKAPFRKSHLETHIVSHSEK 82  
DB 210 KPYIC--ECCGKAFKYSALNTHKRIHTGKPYKC--DKCDKAFIASSTLSKHEIHTGK 265  
QY 83 KPFFCSVCGVGNRSQHLKRHEIHT--KSFKCTFENCQAFYKHQSLRHH-----131  
DB 266 KPYKCECGKAFNQSSTLTGKHIHTGKPYKC--ECCGKAFNQSSTLTGKHIHTGK 323  
QY 132 -----ILSVH-----EKTLTCKQCNKVTRPSKLAQHLKH-----162  
DB 324 YVCECGKAFKYSRILTHKRIHTGKPYKCNKCGKAFIASSTLSRHEFTGKHKHYKCE 383  
QY 163 -----HGGSPAYQCDHPGCFKFNQFQWVLOPHIKOSH---PKLKCPK 201  
DB 384 ECGKAFIWSVLTGKHIHTGKPYKCEE--CGKAFKYSSTLSH--KRSHTGKPYKCE 440  
QY 202 CGKGCVGKGLSSHMLSHDDSTMIKIWTCDYGVGK--FAKKNELVEHYNIHFDGNIPDDL 260  
DB 441 CGKAFVASTLSKHEIHT--TGKKPYKCEC--GKAFNQSSTLTGKHIHTGK 491  
QY 261 LKETEYKLENLLDQGSKLNNLHELET--EKLKVEDEDEDEDS---LDEKRSVRSDSM 315  
DB 492 YKCECGKAFN---QSSSLTGKHIHTGKPYKCECGKAFNQSSTLTGKHIHTREKPY 548  
QY 316 SAORSIKSFASLEGSKSVKLSNSGKKNPCPNNDMFRSREYDLRRLKWH 369  
DB 549 KCECGKAF--HLSTHLTTHKILHTGKPYRC--RECCKAFNHSATLSHKKI 598

RESULT 14  
ABG18386  
ID ABG18386 standard; protein; 632 AA.  
XX  
AC ABG18386;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #18377.

```

XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS82573.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 48745; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 632 AA;
XX
XX Query Match 17.7%; Score 395; DB 4; Length 632;
XX Best Local Similarity 29.7%; Pred. No. 1.9e-22;
XX Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;
XX
XX 23 KKYICTYECGDKAYNRPSSLEOHLRTHSNDRPYKCTVDDCDKAFPRKSHLETHIVSHSEK 82
XX
XX 211 KPYIC--EECGAFKYSALNTHKRIHTGKPYKC--DKCDKAFASSTLSKHEIHTGK 266
XX
XX 83 KPHCSVCGKYNRSOHLKREHITHT--KSFKCTFENCQEAFFKQSLRHH----- 131
XX
XX 267 KPYKECGKAFNQSTLTGKHKIHTGKPYKC--EECGAFNQSTLTGKHKIHTGK 324
XX
XX 132 -----ILSVH-----EKLTCQCKNKFVTRPSKLAQHLKH----- 162
XX
XX 325 YVCEECGKAFKYSRIILTHKRIHTGKPYCKNCKGKAFASSTLSRHEFIHMGKHKYKCE 384
XX
XX 163 -----HGGSPAYOCDPGCFKQKQFQVSWLQFIKQSH---PKLKCPK 201
XX
XX 385 ECGKAFIWSVLTRHVRHTGKPYKCEB--CGKAFKYSSTLSH--KRSHTGKPYKCEB 441

```

```

QY 202 CGKCVGKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIHFGNIPDDL 260
DB 442 CGKAFVASSTLSKHEIHT--TGKPYKCEB--GKAFNQSSSLTKKKI-HTGKFP--- 492
QY 261 LKSTEVKLENLLDQSKLNNLHLELT--EKLKVEDEDEDEDS---LDEKSDVSDSM 315
DB 493 YKCEECGKAFN---QSSSLTKKHKIHTGKPYKCECGKAFNQSSSLTKKKIHTREKPY 549
QY 316 SAQRSIKSFASLEGSKSVSKLISNSGKKINCNNKNCNDRMFSREYDLRLHLKWH 369
DB 550 KCEECGKAF--HLSTHLTTHKIHTGKPYRC--RECGRAFNHSATLSHKKIHT 599

RESULT 15
ABG16953
ID ABG16953 standard; protein; 719 AA.
XX
XX AC ABG16953;
XX
XX DT 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #16944.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX FN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US008631.
XX
XX PR 31-MAR-2000; 2000US-00540217.
XX
XX PR 23-AUG-2000; 2000US-00649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX
XX DR N-PSDB; AAS81140.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX PS Claim 20; SEQ ID NO 47312; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:18:46 ; Search time 44 Seconds  
(without alignments)

2954.398 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229

Sequence: 1 MRSSTKTSISLSSSSSS.....PLVKKARMDLLNETSVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL 25:\*

2: sp\_archaea:\*

3: sp\_bacteria:\*

4: sp\_fungi:\*

5: sp\_human:\*

6: sp\_invertebrate:\*

7: sp\_mammal:\*

8: sp\_mhc:\*

9: sp\_organelle:\*

10: sp\_phase:\*

11: sp\_plant:\*

12: sp\_rodent:\*

13: sp\_virus:\*

14: sp\_vertebrate:\*

15: sp\_unclassified:\*

16: sp\_rvirus:\*

17: sp\_bacteriap:\*

18: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	410.5	18.4	634	11	Q8bi85 mus musculus
2	400	17.9	404	4	Q8n7Q3 homo sapien
3	397.5	17.8	395	4	Q43693 homo sapien
4	392.5	17.6	542	4	Q8iYNO homo sapien
5	392	17.6	555	4	Q96N38 homo sapien
6	392	17.6	644	4	Q8ND40 homo sapien
7	392	17.6	665	4	Q86W65 homo sapien
8	391	17.5	592	4	Q8n8Q4 homo sapien
9	390	17.5	540	4	Q8N211 homo sapien
10	390	17.5	576	4	Q96N22 homo sapien
11	388.5	17.4	501	5	Q9VG72 drosophila
12	388	17.4	568	4	Q8TD23 homo sapien
13	387	17.4	559	11	Q80VH2 mus musculus
14	385.5	17.3	297	11	Q8BU46 mus musculus
15	382	17.1	511	11	Q8VEC1 mus musculus
16	381	17.1	493	11	Q921H6 mus musculus

17	379.5	17.0	710	11	Q91VP4 mus musculus
18	379	17.0	600	11	Q8C300 mus musculus
19	379	17.0	766	11	Q8C4J2 mus musculus
20	378.5	17.0	623	11	Q62510 mus musculus
21	378.5	17.0	914	11	Q8C827 mus musculus
22	378	17.0	393	11	Q92352 mus musculus
23	378	17.0	452	11	Q8BQC2 mus musculus
24	378	17.0	452	11	Q8BPP0 mus musculus
25	378	17.0	613	4	Q8H7R5 homo sapien
26	376	16.9	751	4	Q9TBA9 homo sapien
27	374.5	16.8	710	11	Q61898 mus musculus
28	374	16.8	578	11	Q8BLA2 mus musculus
29	374	16.8	578	11	Q7TWN8 mus musculus
30	374	16.8	641	4	Q86T91 mus musculus
31	374	16.8	641	4	Q86T29 mus musculus
32	374	16.8	650	11	Q91WF9 mus musculus
33	373.5	16.8	517	4	Q8N492 mus musculus
34	373	16.7	400	11	Q8VHT7 mus musculus
35	373	16.7	747	4	Q86XU8 mus musculus
36	372.5	16.7	586	4	Q8NB35 mus musculus
37	372	16.7	637	4	Q86IR2 mus musculus
38	372	16.7	737	11	Q7TSH9 mus musculus
39	369.5	16.6	464	4	Q8N8T4 mus musculus
40	369	16.6	531	4	Q8NE34 mus musculus
41	369	16.6	650	4	Q86XL7 mus musculus
42	368	16.5	488	4	Q9NV05 mus musculus
43	368	16.5	488	4	Q96K00 mus musculus
44	367.5	16.5	462	11	Q8K286 mus musculus
45	366.5	16.4	393	4	Q9H6Z6 mus musculus

## ALIGNMENTS

### RESULT 1

Q8bi85 ID Q8bi85 PRELIMINARY; PRT; 634 AA.

AC Q8bi85;  
DT 01-VAR-2003 (TREMELrel. 23, Created)  
DT 01-VAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Similar to DNA-binding protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cortex;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK043955; BAC31714.1;  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001909; KRAP.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF01352; KRAP; 1.  
DR Pfam; PF00096; zf-C2H2; 16.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR SMART; SM00349; KRAP; 1.  
DR SMART; SM00355; Znf\_C2H2; 16.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 16.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_2; 16.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 16.  
SQ SEQUENCE 634 AA; 73668 MW; AA7DEE33D87508BA CRC64;

Query Match

18.4%; Score 410.5; DB 11; Length 634;

Best Local Similarity 32.8%; Pred. No. 2e-19;  
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

QY 23 KKYICTYEGCDKAYNRPSLLQHLRTHSNDPRPYKCTVDDCDKAFKSHLETHIVSHSEK 82  
Db 323 KPYKC--EECGKAFNSSILTKHILHTGKPKYC--EECGKFSVSLTHTKAIHAE 94  
QY 83 KPHFSGVCGKVNROHLKRHEIHT--KSFCTTENCQBAFYKQSLRHILSVH--EK 138  
Db 379 KPYCNICGKSFNSCTNLTKHTLTHTGKPKYC--KECGKAFPMSSLIHQNLHPGK 436  
QY 140 LTCQCNKVFTRPSKLAQHLKHGGSPAYOCDHGCFKQFQWVSLQFH--IKQSHPKL 197  
Db 437 YKCECDKSFLLKSLRTHQIHTGKPKYC--CGKSFNCTNLTKHTLTHTGKPKY 493  
QY 198 KCPKCGKCGVCKGLSSMLSHDDSTMTKIWTCDYCDVGFPAKNELVHYNIHFHGNIP 257  
Db 494 KCECGKSFHYWSSLKSHQNLHSGE--KPYKCECD-KSFTEKSTLIHQRI-HTG-- 545  
QY 258 DDLKETEVEKLE--NLLDQSKLNLHLETEKLVKVEDEEDSDLSDEKRSVRSDSM 315  
Db 546 -----KKLYKCNICDK--SFTWCASLTKTHK-KHTGK-----PY 577  
QY 316 SAQSIKSF--TAGLESGKSVKLSNSGKKI-NCPNNDMPFSREYDLRHLKWH 369  
Db 578 KCRKCGKSFQPLSTKRGKHIEDKHTGKLYKC--NDCRSYSYHSSFRHOKIH 632

RESULT 2  
Q8N703 PRELIMINARY; PRT; 404 AA.

AC Q8N703  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein FLJ40479.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,  
RA Matsunoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,  
RA Matsunura Y., Moriya S., Chiba E., Morioka H., Onogawa S.,  
RA Kasaiyama S., Satoh N., Matsunawa H., Takahashi B., Kataoka R.,  
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,  
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,  
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,  
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,  
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,  
RA Sugano S., Nagahara K., Masuko Y., Nagai K., Isogai T.,  
RA "NEDO human cDNA sequencing project";  
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RL -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC EMBL; AK097798; BAC05174.1;  
DR GO; GO:0005634; C:nucleus; IEA.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; Znf\_C2H2\_1.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR SMART; SM00355; Znf\_C2H2\_13.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 14.  
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;  
KW Zinc-finger.  
SQ SEQUENCE 404 AA; 45923 MW; B1B35F62F5DF0007 CRC64;

Query Match 17.9%; Score 400; DB 4; Length 404;  
Best Local Similarity 33.9%; Pred. No. 6.3e-19;  
Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;

QY 23 KKYICTYEGCDKAYNRPSLLQHLRTHSNDPRPYKCTVDDCDKAFKSHLETHIVSHSEK 82  
Db 39 KPYKC--EECGKAFNSSILTKHILHTGKPKYC--EECGKFSVSLTHTKAIHAE 94  
QY 83 KPHFSGVCGKVNROHLKRHEIHT--KSFCTTENCQBAFYKQSLRHILSVH--EK 138  
Db 95 KPYKCECGKASNSSSKLMHKKRIHTGKPKYC--EECGKAFSSWSSSLTEH-KRIHAGEK 151  
QY 139 TLTCQCNKVFTRPSKLAQHLKHGGSPAYOCDHGCFKQFQWVSLQFH--IKQSHPK 196  
Db 152 PYKCECGKAFNSSILTKHILHTGKPKYC--GGKAFSKVSTLTHTKAIHAEK 208  
QY 197 LKPKCGKCGVCKGLSSMLSHDDSTMTKIWTCDYCDVGFPAKNELVHYNIHFHGN 255  
Db 209 YKCECGKASNSSSKLMHKKRIH--TGKPKYKCEC--GKAFSSWSSSLTEH-KRIHAGEK 262  
QY 256 IPDILLKETEVEKLE--NLLDQSKLNLHLETEKLVKVEDEEDSDLSDEKRSVRSDS 314  
Db 263 KP---YKCECGKAFSSWSSFTKHEIHAEE-KPYKCECGKGFSTFTLTHTKHIHTGE 318  
QY 315 --MSAQRISKSF--TAGLESGKSVKLSNSGKKI-NCPNNDMPFSREYDLRHLKWH 369  
Db 319 KRYKCECGKAF--SSWSSILTEHILHTGKPKYC--EECGKAFSSWSSSLTRHKKRIH 371

RESULT 3  
O43693 PRELIMINARY; PRT; 395 AA.

AC O43693  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Zinc-finger protein (Fragment).  
CN ZFS-1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testicular tumor;  
RA Ogawa T., Poncellet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,  
RA Udagawa K., Lecocq P., Marine J., Martial J., Hosaka M.,  
RT "Enhanced expression in seminoma of human zinc finger genes located on  
chromosome 19";  
RL Cancer Genet. Cytogenet. 100:36-42(1998).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; D70831; BAA24050.1;  
DR HSP; P08048; 7ZNF.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0008237; F:metalloproteinase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; Znf\_C2H2\_13.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR ProDom; PD000003; Znf\_C2H2\_10.  
DR SMART; SM00355; Znf\_C2H2\_13.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 13.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 13.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.  
FT NON TER 1  
SQ SEQUENCE 395 AA; 45291 MW; D4E06B54DCE0BF9B CRC64;

Query Match 17.8%; Score 397.5; DB 4; Length 395;  
Best Local Similarity 35.9%; Pred. No. 9e-19;  
Matches 97; Conservative 35; Mismatches 87; Indels 51; Gaps 14;

```

QY 23 KYVICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVSHSEK 82
Db 133 KPYKC--KECGKAFNRSTLTTRKHTGKEPKYC--EECGKALTOSSHLTTHKIHTGE 188
QY 83 KPFHSCVCGKGVNSRQHLKHEITHT--KSFKCTFENCQAF-----122
Db 189 KPYKCKKCGKAFNQSAHLTTHVHTGKEPKYC--EKCGKAFNHFHSHLTTHKIHTGKEP 246
QY 123 -----YKHQS--LRHILSVHEKTLCKQCNKVFTEPSPKLAQHLKHHGSPAYQC 171
Db 247 YKCEKCGKAFKHSSTLTTRKHTGKEPKYC--EKCGKAFNHFHSHLTTHKIHTGKEP 305
QY 172 DHPGCFNFQTSVLQPHIKQSHPK---LKPCKGCGKGVCKGLSSHMLSHDDSTMIKIW 228
Db 306 EE--CGKAFNQSNLTRH--KKSHTPEPKYKCECGKALNRPSTLTTHKIHTGKEP 359
QY 229 TCDYCDVGK-PAKQVELVEHYNIFHDGNIP 257
Db 360 KCEEC--GKAFNQSKLTTHKHI--HTGKEP 386

RESULT 4
Q81YNO ID Q81YNO PRELIMINARY; PRT; 542 AA.
AC Q81YNO;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similar to zinc finger protein 208.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Testis;
RA Strauberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC035579; AAH35579.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 11.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 11.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
SQ SEQUENCE 542 AA; 62756 MW; 395709FE7E094C9B CRC64;

Query Match 17.6%; Score 392.5; DB 4; Length 542;
Best Local Similarity 30.0%; Pred. No. 2.7e-18;
Matches 115; Conservative 39; Mismatches 132; Indels 97; Gaps 19;

QY 8 KSISLISSSSSR----PKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCD 63
Db 212 KSFNLLLTQKRPHTITENSVC--KDCGKAFNFWSTLTTRHRIHTGKEPKYC--EECG 267
QY 64 KAPFRKSHLETHIVSHSEKPFHSCVCGKGVNSRQHLKHEITHT--KSFKCTFENCQEA 121
Db 268 KAFNRSSHLTTHKIHTGKEPKYKCECGKAFNRSSHLTTHKIHTGVKPYKCT--EECGA 325
QY 122 FYKHQS--LRHILSVHEKTLCKQCNKVFTEPSPKLAQHLKHHGSPAYQCDHPGCFK 180
Db 326 FNRSSHLTTHRIHTGKEPKYKCECGKAFNQSNLTTHKIHTGKEP--YKCEE--CGKAF 382
QY 181 QTSVSLQPHIKQSHPK---LKPCKGCGKGVCKGLSSHMLSHDDSTMIKIHTCDYCDVGK 237

```

```

Db 383 YRFSYLIKH-KTSHTGKPYKCECGKGFNWSSALTTRKRIH-----TGKPYKCEC--GK 436
QY 238 -PAKQVELVEHYNIFHDGNIPDDLLKETEYVKLENLLDQGSKLNNLHELETEKLKVEDE 296
Db 437 AFNENSLTTH-KMIHTGKEP---YKDECGKAFN-----467
QY 297 EDEEDSLDKRSVDRSDMSAQRISIKSTASLEGSKSVSKLISNGKINCPKNCNDMP 356
Db 468 -----RSSQLTAH-----KMIHTGKEPKYC--EECGKAF 494
QY 357 SREYDLRSH-----LKWHD 370
Db 495 NRSSTLTTRKHTGKEKSYKWE 517

RESULT 5
Q96N38 ID Q96N38 PRELIMINARY; PRT; 555 AA.
AC Q96N38;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ31444.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RL "NEO human cDNA sequencing project.";
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL; AK056006; BAB71072.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf-C2H2; 13.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW Zinc-finger.
SQ SEQUENCE 555 AA; 64025 MW; 75C246D4820FB920 CRC64;

Query Match 17.6%; Score 392; DB 4; Length 555;
Best Local Similarity 32.5%; Pred. No. 3e-18;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17;

QY 21 RPKKYCTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVSHS 80
Db 169 RENSVC--EECDKVPKRPSTLTTRKRVHTGKEPKYC--EECGKAFKHSSTLTTHKMIHT 224
QY 81 EKPFHSCVCGKGVNSRQHLKHEITHT--KSFKCTFENCQAFYKHQSL--RRHILSVHE 137
Db 225 GEKPYKCECGKAFYHSHSLTTRKVIHTGKEPKYC--EECGKAFNRPSTLTTHKFIHVKE 282
QY 138 KTUTCQCNKVFTRPSKLAQHLKHHGSPAYQCQDHGCFKQFQTSVSLQFH--IKQSHP 195

```

```

Db 293 KPYKCECDKAFNPSYLTTHKHI-IHSGEKSYKCEQ--CGKGFNWSSTLTTHKRIHTGK 339
QY 196 KLKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNI PHDGN 255
Db 340 PYKCECGKAFNVSSTLTTHKHI--TGEKPYKCEC--GKAFNHSSKLTTHKRIHTG 394
QY 256 IPDDLKETEYKLENDLQSGKLNHLHELET-EKL-KVEDEDEDESD--EKRSV 310
Db 395 KP---YKCECGKAFN---QSSNLTTHKHIHTGEKLYKCECGKAFNRSNLTTHKRIHT 448
QY 311 RSDMSAQRISKFTASLEGSKSVKLSNSGKKINCPKNCNDRMFSREYDLRRHLK 367
Db 449 GEKPYKCECGKAFNRSNLTG--HNIHTGKSYKC--EBCGKAFNQSSTLTTHKRIHT 501

RESULT 6
QND40
ID Q8ND40 PRELIMINARY; PRT; 644 AA.
AC Q8ND40
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKZP5470168.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RA Bloecher H., Boscher M., Brandt P., Neues H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AL834415; CAD39077.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf-C2H2; 13.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW Zinc-finger.
FT NON_TER
SQ SEQUENCE 644 AA; 74012 MW; A3CBF2437279B863 CRC64;

Query Match 17.6%; Score 392; DB 4; Length 644;
Best Local Similarity 32.5%; Pred. No. 3.5e-18;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17;

QY 21 RPKYICTYEGCDKAYNRPSILLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVGHS 80
Db 289 RENSQC--EEDKVFKEFSTLTTHKRVHTGKFPKC--EBCGKAFKHSSTLTTHKMIHT 344
QY 81 EKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFCTTFENCQEAIFYKHQS-LRHHLSVHE 137
Db 345 GEKPYRCECGKAFYHSSHTLTHKVIHTGKFPKC--EBCGKAFNPSALTTHKTHIHYKE 402
QY 138 KTLTCKCNKVFTRPSKLAQHLKHLKGGSPPAYOCDPHGCFCNFTQWSVLQPH--IKQSH 195
Db 403 KPYKCECDKAFNPSYLTTHKHI-IHSGEKSYKCEQ--CGKGFNWSSTLTTHKRIHTGK 459
QY 196 KLKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNI PHDGN 255
Db 409 KPYKCECDKAFNPSYLTTHKHI-IHSGEKSYKCEQ--CGKGFNWSSTLTTHKRIHTGK 459
QY 196 KLKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNI PHDGN 255

```

```

Db 460 PYKCECGKAFNVSSTLTTHKMIH--TGEKPYKCEC--GKAFNHSSKLTTHKRIHTG 514
QY 256 IPDDLKETEYKLENDLQSGKLNHLHELET-EKL-KVEDEDEDESD--EKRSV 310
Db 515 KP---YKCECGKAFN---QSSNLTTHKHIHTGEKLYKCECGKAFNRSNLTTHKRIHT 568
QY 311 RSDMSAQRISKFTASLEGSKSVKLSNSGKKINCPKNCNDRMFSREYDLRRHLK 367
Db 569 GEKPYKCECGKAFNRSNLTG--HNIHTGKSYKC--EBCGKAFNQSSTLTTHKRIHT 621

RESULT 7
Q86W65
ID Q86W65 PRELIMINARY; PRT; 665 AA.
AC Q86W65
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to zinc finger protein 91 (HPF7, HTF10) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050468; AAH50468.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 13.
DR PRINTS; PR00048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 11.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
FT NON_TER
SQ SEQUENCE 665 AA; 76467 MW; 1ABDA1E4E00249DF CRC64;

Query Match 17.6%; Score 392; DB 4; Length 665;
Best Local Similarity 32.5%; Pred. No. 3.7e-18;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17;

QY 21 RPKYICTYEGCDKAYNRPSILLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVGHS 80
Db 279 RENSQC--EEDKVFKEFSTLTTHKRVHTGKFPKC--EBCGKAFKHSSTLTTHKMIHT 334
QY 81 EKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFCTTFENCQEAIFYKHQS-LRHHLSVHE 137
Db 335 GEKPYRCECGKAFYHSSHTLTHKVIHTGKFPKC--EBCGKAFNPSALTTHKTHIHYKE 392
QY 138 KTLTCKCNKVFTRPSKLAQHLKHLKGGSPPAYOCDPHGCFCNFTQWSVLQPH--IKQSH 195
Db 393 KPYKCECDKAFNPSYLTTHKHI-IHSGEKSYKCEQ--CGKGFNWSSTLTTHKRIHTGK 449
QY 196 KLKCPKCGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNI PHDGN 255
Db 450 PYKCECGKAFNVSSTLTTHKMIH--TGEKPYKCEC--GKAFNHSSKLTTHKRIHTG 504
QY 256 IPDDLKETEYKLENDLQSGKLNHLHELET-EKL-KVEDEDEDESD--EKRSV 310
Db 505 KP---YKCECGKAFN---QSSNLTTHKHIHTGEKLYKCECGKAFNRSNLTTHKRIHT 558
QY 311 RSDMSAQRISKFTASLEGSKSVKLSNSGKKINCPKNCNDRMFSREYDLRRHLK 367

```

Db 559 GKPYKCECGKAFNRSSNLTK--HNIIHTGKSYKC--EBCGKAFNQSSLTGKRX 611

RESULT 8

Q8N8Q4 PRELIMINARY; PRT; 592 AA.

AC Q8N8Q4; (TREMELrel. 22, Created)

DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein FLJ39023.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Nimmiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,

RA Furuwa T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,

RA Katsuta N., Sato H., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,

RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,

RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,

RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,

RT "NEO human cDNA sequencing project."

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; AK096342; BAC04764.1; -.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0008237; F:metallopeptidase activity; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR GO; GO:0008270; F:zinc ion binding; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001309; KRAB.

DR InterPro; IPR007087; Znf\_C2H2\_BS.

DR InterPro; IPR007086; Znf\_C2H2\_sub.

DR Pfam; PF01352; KRAB; 1.

DR Pfam; PF00096; Zf-C2H2; 15.

DR PRINTS; PD00048; ZINC\_FINGER.

DR ProDom; PD000003; Znf\_C2H2; 5.

DR SMART; SM00349; KRAB; 1.

DR SMART; SM00355; Znf\_C2H2; 16.

DR PROSITE; PS0805; KRAB; 1.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 15.

DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 16.

DR PROSITE; PS00142; ZINC\_PROTEASE; 4.

KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc; Zinc-finger.

SK SEQUENCE 592 AA; 67920 MW; D116FB9366EB4BDE CRC64;

Query Match 17.5%; Score 391; DB 4; Length 592;

Best Local Similarity 31.1%; Pred. No. 3.8e-18;

Matches 120; Conservative 41; Mismatches 161; Indels 64; Gaps 19;

QY 23 KKYICTYGGCKAYNRPRLLEQLRTHNSNDPRPKYCTVDDCDKAFRRKSHLETHIVSHSEK 82

Db 199 KPYIC--EBCGKAFKYSALNTHKRIHTGKPYKC--DKCDKAFIASSTLSKHEIHTGK 254

QY 83 KPFHSCVCGKGNRSQHLKHEIHT--KSPKCYFENCQAFYKHQSLRH-- 131

Db 255 KPYKCECGKAFNQSSLTGKRIHTGKPYKC--EBCGKAFNQSSLTGKRIHTGK 312

QY 132 -----ILSVH-----EKLTLCKQCNKVFTRPSKLAQHLKHGSGPAYQC 171

Db 313 YVCBCGKAFKYSRLTLTKRIHTGKPYKCNKCGKAFIASSTLSRHEFIHM-GKKYK 371

QY 172 DHPGGCFKMFQVSWLQFH--IKQSHPLKCPKCGKCGVKGKGLSSMLSHDDSTMTKIWT 229

Db 372 EE--CGKAFIWSVLTTRHVRHTGKPYKCECGKAFIASSTLSKHEIHT--TGKKPYK 426

QY 230 CDYCDVGK-FAKKNELVHYNIIFHDGNIIPDLLKETEYKLENNLLDQSGSLNNLHELET- 287

Db 427 CEEC--GKAFNQSSLTGKRIHTGKPYKC--YKCECGKAFN---QSSSLTKHKIHTG 477

QY 288 -EKLKVEDEDEDEDS---LDEKESDVRSQMSAQRISIKSFTASLEGSKSVKLSNSGK 343

Db 478 EKPYKCECGKAFNQSSLTGKRIHTGKPYKCECGKAF--HLSTHLTHKILHTGK 535

QY 344 KINCPCNNCDRMFSREYDLRRHLKWH 369

Db 536 PYRC--RECCKAFNHSATLSSHKKI 559

RESULT 9

Q8N211 PRELIMINARY; PRT; 540 AA.

AC Q8N211;

DT 01-OCT-2002 (TREMELrel. 22, Created)

DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein FLJ36350.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,

RA Arima M., Mutsaers K., Yuuki H., Hara H., Sugiyama T., Irie R.,

RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,

RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,

RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,

RT "NEO human cDNA sequencing project."

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; AK093659; BAC04216.1; -.

DR PIR; P42075; P42075.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001909; KRAB.

DR InterPro; IPR007087; Znf\_C2H2.

DR InterPro; IPR007086; Znf\_C2H2\_sub.

DR Pfam; PF01352; KRAB; 1.

DR Pfam; PF00096; Zf-C2H2; 12.

DR PRINTS; PD00048; ZINC\_FINGER.

DR ProDom; PD000003; Znf\_C2H2; 1.

DR SMART; SM00349; KRAB; 1.

DR SMART; SM00355; Znf\_C2H2; 12.

DR PROSITE; PS0805; KRAB; 1.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 11.

DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 13.

KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc; Zinc-finger.

SK SEQUENCE 540 AA; 62985 MW; 84D7D2EC15A7002C CRC64;

Query Match 17.5%; Score 390; DB 4; Length 540;

Best Local Similarity 30.0%; Pred. No. 4e-18;

Matches 107; Conservative 61; Mismatches 155; Indels 34; Gaps 15;

QY 23 KKYICTYGGCKAYNRPRLLEQLRTHNSNDPRPKYCTVDDCDKAFRRKSHLETHIVSHSEK 82

Db 198 KLYKC--QECDFTFNQSNLTGKRIHTGKPYKCECGKAFNQSSLTGKRIHTGK 253

QY 83 KPFHSCVCGKGNRSQHLKHEIHTKSKFCTPENCQAFYKHQSL-RHILSVHEKTLT 141

Db 254 KPYKCECGKAFNQSSLTGKRIHTGKPYKCECGKAFNQSSLTGKRIHTGKPYK 313

QY 142 CKQCNKVFTRPSKLAQHLKHGSGPAYQCDHPCGCFNFTQWSVLOFH--IKQSHPLK 199

Db 314 CEECGAFNRSSKJTEHKNHTGEOP-YKCEE--CGKAFNRSSNLTHERKIHTTEPKYC 370  
 QY 200 PKCGKGVCKGLSSHMLSHDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIHFDGNIPD 258  
 Db 371 KECGKAFHSSALITTHKRIH---TCEPKYKCEEC--GKAFNRSSKJTEHKKL-HTCK---421  
 QY 259 DLLKTEVYKLENLLDQSKLNNLHELETKLVEDE-----EDEDSLDKXSDVRSD 313  
 Db 422 ---KPYKECGKAFIQSGKJTEHKKHSHGPIPKYCECGKAFHSSSLTTHKRIHTGK 478  
 QY 314 SMSAQRISKFTASLEGSKSVKLSNSGKK-INCPKNNCDRMFSREYDLRRLKWH 369  
 Db 479 PKYCECGKAFSRS---SKLTEHKIHTGKPKYKCEC--CDKAFNQSANLTTHKKLIH 530

RESULT 10  
 Q96N22 ID Q96N22 PRELIMINARY; PRT; 576 AA.  
 AC Q96N22;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein FLJ1526.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,  
 RT "NEO human cDNA sequencing project".  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL; AK056088; BABY1090.1; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Dfam; PF00096; zf-C2H2; 15.  
 DR PRINTS; PR00048; ZINC\_FINGER.  
 DR ProDom; PD000003; Znf\_C2H2; 2.  
 DR SMART; SM00355; Znf\_C2H2; 16.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 10.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 16.  
 KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 576 AA; 66620 MW; 4396672D34BF99CD CRC64;

Query Match 17.5%; Score 390; DB 4; Length 576;  
 Best Local Similarity 28.6%; Pred. No. 4.3e-18;  
 Matches 119; Conservative 57; Mismatches 168; Indels 72; Gaps 17;

QY 3 ESDTKSTSSLSRSSSRPKYICTYEGCDKAYNRPSLLEQLHRTSHSDRPYKCTVDDC 62  
 Db 217 ECDKAFNQSSLTTHKRIHTREKLEKNEKCGKAFNQSSSLTTHKRIHTGKPKYC--EEC 274  
 QY 63 DKAFPKSHLETHIVSHSEKPPFHCSCGKGVNSRQHLKHEIHT--KSFKCTFNCQE 120  
 Db 275 GKAFNQSSSLTTHKRIHTGKPKYCECGKAFNQSSSLTTHKRIHTGKPKYC--EECGK 332  
 QY 121 APYKQSL--RHILSVHRTKTCQKNVFRTPPSKLAQHLKXHGSGSPAYQCQHPGCFKN 179  
 Db 333 AFNKKSSSLTTHKRIHTGKPKYCECGKAFNQSSSLTTHKRIHTGKPKYC--CGKA 389  
 QY 180 FQTVSLVQFH--IKOSHPLKCPKCGKGVCKGLSSHMLSHDSTMIKIWTCDYCDVGK 237  
 Db 390 FNFQNSLTHKRIHTGKPKYCECGKAFNQSSSLTTHKRIHT--TCEKSYKCEEC--GK 444

## RESULT 11

Q9VG72 ID Q9VG72 PRELIMINARY; PRT; 501 AA.  
 AC Q9VG72;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE CG5245 protein.  
 GN CG5245.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=Berkeley;  
 RC MEDLINE=20194006; PubMed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Adamantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Fertaz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Testor C., Turner B., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003696; AAF54813.1; -.  
 DR HSSP; P08047; 1SP2.  
 DR FlyBase; FBgn0036047; CG5245.

```
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 15.
DR ProDom; PD00003; Znf_C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 15.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 15.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 15.
DR Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 501 AA; 58790 MW; BBAE04740D41C43F CRC64;

Query Match 17.4%; Score 388.5; DB 5; Length 501;
Best Local Similarity 29.7%; Pred. No. 4.6e-18;
Matches 109; Conservative 63; Mismatches 142; Indels 53; Gaps 19;

QY 21 RPKYICTYECGDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFKRHLETHIVSHS 80
DB 135 RPKH--CSH--CSKTFQNSSLKQHLHEHTGERPFKCT--QCSTSFARKSHLQVHLRTHS 188
QY 81 EKKPSPHSCVCGKGVNSRQHLKHEITH--TKSFCTFENCQEAIFYKHQSLRHLSVHEK 138
DB 189 EERPECTHCBAKFNNSHLEHRLTHQEARPFKCS--HCKSKFLRSILQKHLTHAER 246
QY 139 TLTCQCNKVFTRPSKLAQKHLKHGGSPAYQCDHPGCFKNTQVSWVLQPHIKQSHPK-- 196
DB 247 SFKCTQCPKTFLODSLQIH--LRVHAGEDDPKCPH--CSETFARNRSLQLHLEHACKEP 303
QY 197 LKCPKCGKGVGKGLSHMLSHDSTMKIWTCDYCDVGKFAKNELVEHYNIFFDQNI 256
DB 304 LKCSQCSATFAMRSLRVHVLH--TRERQYKCAECS--KSFFPKSHLVEHQV--HTGER 358
QY 257 P-----DDLKTEVKLENLLDQ--GSKNLNLELETE-----KLKVEDEDEDEDSL 303
DB 359 PFKCTHCFKDKRTHLR--VHMLDHIGKVPKSCYCKEFLSSQLLVLQHTGKNQF 416
QY 304 DEKSDVRSMSQORSIKSTASLEGSKSVKLSINSKG--KINCPKNKCDRPMFSRYDL 362
DB 417 E-----CPHCSKSYTTS---STLHMLRTHTGELPPFKC--SHCSKLFARSAEH 459
QY 363 RRLHLKWH 369
DB 460 QEHLATH 466

RESULT 12
Q8TD23 PRELIMINARY; PRT; 568 AA.
AC Q8TD23;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TRAF6-binding zinc finger protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21864204; PubMed=11751921;
RA Shin J.N., Kim I., Lee J.S., Koh G.Y., Lee Z.H., Kim H.H.;
RT "A Novel Zinc Finger Protein That Inhibits Osteoclastogenesis and the
RT Function of Tumor Necrosis Factor Receptor-associated Factor 6.";
RL J. Biol. Chem. 277:8346-8353 (2002).
DR EMBL; AY044432; AAK95822.1; -.
DR FIR; F42075; F42075.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001809; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 13.
DR PRINTS; PRO0048; ZINC_FINGER.
DR ProDom; PD000003; Znf_C2H2; 1.
```

```
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 14.
DR Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 568 AA; 66222 MW; 3BBC819822B89940 CRC64;

Query Match 17.4%; Score 388; DB 4; Length 568;
Best Local Similarity 28.5%; Pred. No. 5.7e-18;
Matches 105; Conservative 62; Mismatches 167; Indels 34; Gaps 13;

QY 15 SSSSSRPKPYIC----TYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFKRS 70
DB 212 SKLTKHGIYCEKLYKQCECDRTFNQFNLSNTEYKDYAREKPYK--EEGKAFNQSS 269
QY 71 HLETHIVSEKPPHSCVCGKGVNSRQHLKHEITH--TKSFCTFENCQEAIFYKHQSL-R 129
DB 270 HLTTHKIHTGEKPYKCECGKASQFNLSLTHKXIHTGEQYICECGKAFQSTLT 329
QY 130 HILSVHEKTLTCQCNKVFTRPSKLAQKHLKHGGSPAYQCDHPGCFKNTQVSWVLQPH 189
DB 330 HKRIHTGEKPYKCECGKAFNRSSKLTENKHTHTGEQ--YKCEE--CGKAFNRSSNLTEH 386
QY 190 --IKQSHPLKCPKCGKGVGKGLSHMLSHDSTMKIWTCDYCDVGKFAKNELVEH 247
DB 387 RKIHTEEKPYKCECGKAFKHSSALTTHKRIH--TGEKPYKCEC--GKAFNRSSKLTE 441
QY 248 YNIFHDGNIPTDLLKTEVKLENLLDQSKNLNLELETEKLVVEDE-----EDEDS 302
DB 442 YKLLHTGK-----KPYKCECGKAFIOSSKLTENKHTHTGEQYICECGKAFKHSS 495
QY 303 LDEKSDVRSMSQORSIKSTASLEGSKSVKLSINSKGK--INCPKNKCDRPMFSREYD 361
DB 496 TTHKRLHTGEKPYKCECSKAFSRS---SKLTEHKIHTHTGEKPYECER--CDKAFNQSAN 550
QY 362 LRRHLKWH 369
DB 551 LTKHKKIH 558

RESULT 13
Q80VH2 PRELIMINARY; PRT; 559 AA.
AC Q80VH2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to zinc finger protein 97.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FVB/N; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049144; AAH49144.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001809; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 15.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 15.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 15.
DR PRINTS; PSS0157; ZINC_FINGER_C2H2_2; 15.
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 15.
SQ SEQUENCE 559 AA; 66355 MW; F456D3FB56B86C39 CRC64;
```

Query Match 17.4%; Score 387; DB 11; Length 559;  
Best Local Similarity 28.7%; Pred. No. 6.6e-18;  
Matches 109; Conservative 55; Mismatches 136; Indels 80; Gaps 17;

QY 23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVSHSEK 82  
Db 187 KPYKCNQ--CDKAFSQHSTLQTHRRTHTGKPYKC--DQCDKAFSEKSLQTHRRTHTGE 242

QY 83 KPHCSVCGKGVNSROHLKRHEITH--KSPKC-----TFEN----- 117  
Db 243 KPYKCNQCDKAFSQSHLHRRTHTGKPLKCNCEDETFNSHNLQTHRRTHTGKPYK 302

QY 118 ---CQAFYKHQSLRHHLS-VHEKTLTCQCNKVTRPSKLAQHLKHGGSAYQODH 173  
Db 303 CNQCDKAFSQHSTLQTHRRTHTGKPYKCNQCDKAFSRHSTLQTHRRTHTGKPYK 361

QY 174 PGCFKQFQWVSLQFHIKQSH---PKLKPCKGCGVGGKGLSSHMLSHDDSTMIKWTC 230  
Db 362 --CDKAFSQYSHLH--RRTHTGKPFKCNQCNKAFSQYSHLHRRTH--TGKPYK 415

QY 231 DYCDVGKFAKBNELVHNIHFHGNIPDDLKETEYKLENLLDQSKLNNLHELETEKL 290  
Db 416 NQCD--KTSNHSSTLQTHRRTHTGE-----KPYKCNQCDKAFSRHSTLQTHRRTHTGK 467

QY 291 KVEDEDEDSLDEKRSVRSMSAQRISIKFTASLEGSKSVSKLINSKGI--NCPK 349  
Db 468 PFKCNQCDK-----AFSQKSLQKHRIHT-----GKLYKC-- 499

QY 350 NNCDRMFSREYDLRRHLKWH 369  
Db 500 NECDKAFSQHSTLQTHRRTH 519

RESULT 14

Q8BJ46

ID Q8BJ46 PRELIMINARY; PRT; 297 AA.

AC Q8BJ46, 2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE MS2F33 homolog.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

DR EMBL; AK032220; BAC27766.1; -

DR InterPro; IPR007087; Znf\_C2H2.

DR InterPro; IPR007086; Znf\_C2H2\_sub.

DR Pfam; PF00096; zf-C2H2; 8.

DR PRINTS; PR00048; ZINC\_FINGER.

DR SMART; SM00355; Znf\_C2H2; 8.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.

DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 8.

SQ SEQUENCE 297 AA; 34566 MW; 2B94C27CB2635EB CRC64;

Query Match 17.3%; Score 385.5; DB 11; Length 297;  
Best Local Similarity 38.1%; Pred. No. 4.2e-18;  
Matches 93; Conservative 36; Mismatches 82; Indels 33; Gaps 14;

QY 23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIVSHSEK 82  
Db 69 KPYKCNQ--CDKAFSRHSTLQTHRRTHTGKPYKC--NQCQKAPTQSHLKHIVHVTGE 124

QY 83 KPHCSVCGKGVNSROHLKRHEITH--KSPKCTFENCQAFYKHQSLRHHLSVHEKTL 140

Db 125 KPYKCNQCGKAFACHNKLQKHERHTHTGKPYKC--DQCNKAF-----VYESYLQVHKTH 177

QY 141 T-----CKCNKVTRPSKLAQHLKHGGSAYQODHPCGCFKFTQWVSLQFHIKQSH 194

Db 178 TGEKPYKCNCGKAFAPARHSHLKVHKIHTGKEP-YKCNQ--CGKALAYHSTLQVH-QRTH 233

QY 195 PKLK--CPKCGKCGVGGKGLSSHMLSHDDSTMIKWTCDYCDVGK-FAKNHELVEHYNI 250

Db 234 TGEKPYCECGKAFANQSYFQVHKRIH---TGKPYKCDQC--GKAFVGSGLKRHERV 288

QY 251 PHDG 254

Db 289 -HTG 291

RESULT 15

Q8VECI

ID Q8VECI PRELIMINARY; PRT; 511 AA.

AC Q8VECI, 2003 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Similar to zinc finger protein 40.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC019219; AAH19219.1; -

DR InterPro; IPR007087; Znf\_C2H2.

DR Pfam; PF00096; zf-C2H2; 12.

DR SMART; SM00355; Znf\_C2H2; 12.

DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.

DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 12.

KW Metal-binding; Zinc; Zinc-finger.

SQ SEQUENCE 511 AA; 59582 MW; 11C75B5E38F2DB6F CRC64;

Query Match 17.1%; Score 382; DB 11; Length 511;  
Best Local Similarity 31.1%; Pred. No. 1.3e-17;  
Matches 110; Conservative 47; Mismatches 143; Indels 54; Gaps 17;

QY 18 SSSRPKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAPFRKSHLETHIV 77

Db 138 TNNREKTYKCS--ECDKCFKKKLRQHLRHTGKPYKCS--ECDKCFQKDLRIHQH 193

QY 78 SHSEKKPFCSCVCGKGVNSROHLKRHEITH--KSPKCTFENCQAFYKHQSLRHHLSV 135

Db 194 VHTGKPYKCSKCDKGTGFTSKRLNHQVHTGKPYKCS--ECDKSPSQGNLSIH-LRI 250

QY 136 H--EKLTLCKCNKVTRPSKLAQHLKHGGSAYQODHPCGCFKFTQWVSLQFHIK 191

Db 251 HTGKPYKCSKCDKCFYKSGLRSHQI--LHTGEKAYKCE--CGKCFTHKGLRRHQSH 307

QY 192 QSHPKLKPCKGCGVGGKGLSSHMLSHDDSTMIKWTCDYCDVGKPAKKNELVEHYNIF 251

Db 308 MGEKPYKCSKCDKGTGFTQKRLIIHQIHTGKGYKCSKCDK--FIKSDLSIHQRI- 362

QY 252 HDGNIPDDLKETEYKLENLLDQSKLNNLHELETEKLKVEDEDEDSLDEKRSVDR 311

Db 363 HTGE-----KPYCEGCKGFTSKRL-NHQ-----RVHTGEK----- 395

QY 312 SDSMSAQRISIKFTASLEGSKSVSKLINSKGIKNCNCDRMPFSREYDLRRH 365

Db 396 --SYKCSKCDKFTQ--QGNLRHLRHTGKPYKC--SECDKCFYKSGLRSH 443

Search completed: May 5, 2004, 15:22:52  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:15:26 ; Search time 17 seconds  
(without alignments)  
1261.936 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229

Sequence: 1 MSEDFTKSISSLISSSSSS.....PLVKKARMDLLNETSVISR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	837.5	37.6	429	1 TF3A_YEAST	P19933 saccharomyc
2	495	22.2	374	1 TF3A_SCHRO	Q9ut15 schizosacch
3	400	17.9	803	1 ZN43_HUMAN	P17038 homo sapien
4	398	17.9	574	1 Z492_HUMAN	Q9p255 homo sapien
5	397.5	17.8	423	1 TF3A_HUMAN	Q92664 homo sapien
6	396.5	17.8	595	1 ZN85_HUMAN	Q93923 homo sapien
7	395	17.7	616	1 ZN93_HUMAN	P35789 homo sapien
8	393.5	17.7	1191	1 ZN91_HUMAN	Q05481 homo sapien
9	387.5	17.4	839	1 Z347_HUMAN	Q068e7 homo sapien
10	385	17.3	576	1 Z431_HUMAN	Q8tf32 homo sapien
11	383	17.2	913	1 Z228_HUMAN	Q9ufu3 homo sapien
12	376	16.9	570	1 Z430_HUMAN	Q9h8g1 homo sapien
13	376	16.9	751	1 Z184_HUMAN	Q99676 homo sapien
14	375.5	16.8	1350	1 XF1N_XENLA	P08045 xenopus lae
15	374	16.8	535	1 Z257_HUMAN	Q9y2q1 homo sapien
16	373	16.7	470	1 Z436_HUMAN	Q9c0f3 homo sapien
17	373	16.7	738	1 ZN94_HUMAN	P15523 homo sapien
18	372	16.7	821	1 ZN41_HUMAN	P18184 homo sapien
19	372	16.7	393	1 ZF46_MOUSE	Q03309 mus musculu
20	367	16.5	818	1 XE18_HUMAN	Q9hcg1 homo sapien
21	367	16.3	428	1 ZN93_HUMAN	P15522 homo sapien
22	363.5	16.3	898	1 Z071_XENLA	P18751 xenopus lae
23	361	16.2	803	1 Z226_HUMAN	Q9nyl6 homo sapien
24	360.5	16.2	630	1 PRD5_HUMAN	Q9nqk1 homo sapien
25	360	16.2	588	1 Z429_HUMAN	Q06v71 homo sapien
26	359	16.1	1167	1 Z208_HUMAN	Q83345 homo sapien
27	357.5	16.0	739	1 YD49_HUMAN	Q9p2j6 homo sapien
28	355	15.9	538	1 Z155_HUMAN	Q12901 homo sapien
29	354	15.9	498	1 Z271_HUMAN	Q14591 homo sapien
30	354	15.9	698	1 Z234_HUMAN	Q14588 homo sapien
31	353	15.8	604	1 Z305_HUMAN	O43309 homo sapien
32	351	15.7	474	1 Z141_HUMAN	Q15928 homo sapien
33	350.5	15.7	783	1 ZF25_HUMAN	Q9ut15 homo sapien

ALIGNMENTS

RESULT 1

ID	TF3A_YEAST	STANDARD;	PRT;	429 AA.
AC	P39933;			
DT	01-FEB-1995 (rel. 31, Created)			
DT	01-FEB-1995 (rel. 31, Last sequence update)			
DT	10-OCT-2003 (rel. 42, Last annotation update)			
DE	Transcription factor Iiia (TFIIIA).			
GN	TFI2 OR PFI1 OR TFIIA OR YPR186C OR P9677.9.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92147684; PubMed=1737784;			
RA	Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,			
RA	Segall J.			
RT	"The deduced sequence of the transcription factor TFIIA from			
RT	Saccharomyces cerevisiae reveals extensive divergence from Xenopus			
RT	TFIIIA."			
RL	J. Biol. Chem. 267:3282-3288(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92237295; PubMed=1570325;			
RA	Woychik N.A., Young R.A.			
RT	"Genes encoding transcription factor Iiia and the RNA polymerase			
RT	common subunit RPB6 are divergently transcribed in Saccharomycetes			
RT	cerevisiae."			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288c / AB972;			
RX	MEDLINE=97133271; PubMed=9169875;			
RA	Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,			
RA	Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,			
RA	Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,			
RA	Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,			
RA	Dierich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,			
RA	Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,			
RA	Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,			
RA	Hunkle-Smith S., Hyman R., Johnston M., Kaiman S., Kleine K.,			
RA	Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,			
RA	Marathe R., Messinguy F., Mewes H.-W., Mirtipati S., Moestl D.,			
RA	Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,			
RA	Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,			
RA	Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,			
RA	Urrestarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,			
RA	Walsh S.V., Wambutt K., Wang Y., Wedler E., Wedler H., Winnett E.,			
RA	Zhong W.W., Zollner A., Vo D.H., Hani J.			
RT	"The nucleotide sequence of Saccharomycetes cerevisiae chromosome XVI."			
RL	Nature 387:103-105(1997).			
CC	-!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF			
CC	APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR			
CC	CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO			
CC	BINDS THE TRANSCRIBED 5S RNA S.			

P17020	homo sapien
Q9uk10	homo sapien
Q9nqz8	homo sapien
P18747	xenopus lae
P17030	homo sapien
Q9ul59	homo sapien
P16374	mus musculus
Q9uk13	homo sapien
O14709	homo sapien
Q9uid9	homo sapien
P34695	rana pipien
P17027	homo sapien



DR PROSITE; PS00028; ZINC FINGER C2H2\_1; 8.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 9.  
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 RN RNA-binding; Repeat; Nuclear protein.  
 FT ZN\_FING 23 47  
 FT ZN\_FING 53 77  
 FT ZN\_FING 83 107  
 FT ZN\_FING 113 138  
 FT ZN\_FING 144 169  
 FT ZN\_FING 204 226  
 FT ZN\_FING 236 261  
 FT ZN\_FING 267 291  
 FT ZN\_FING 349 374  
 SQ SEQUENCE 374 AA; 43851 MW; 7469C701PFF08FF6 CRC64;  
 Query Match 22.2%; Score 495; DB 1; Length 374;  
 Best Local Similarity 32.8%; Pred. No. 1.6e-22;  
 Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;  
 QY 15 SSSSSSRPKYICYEGCDKAYNPSSLLEQHLRTHSNDRPKYCTVDDCDKAFPKSHLET 74  
 DB 13 SKNLSAKKIFHCPEYEGCKYKSPSLLEQHLRTHSNDRPKYCTVDDCDKAFPKSHLKI 72  
 QY 75 HIVSHSEKKPHCSV--CGKGVNSRHLKHEITH--TKSPKCTFENCQAFYKHQSLRH 130  
 DB 73 HKRCHTNVPSCHYDGDCAQFYQQLERHIEVHRPKPYACTWEGDCDFSKHQQLRS 132  
 QY 131 HILSVHEKTL----TCKQCNKVFTRPSKLAQKLKHGGSPAYOCDPGC--FKNFQWTS 184  
 DB 133 HISACHTLLFPYCTYQDCELRFATKQKLQNVNRAHEKIISYSCPHESCVCHEGFEKWS 192  
 QY 185 VLQHIKOSHPLKCPKCKGKCVKGLSSHMLSHDSTMIKWTCDYCDV----GKFAK 240  
 DB 193 QLQNHIREAHVP--SCSICGKQFQTAHLRHVVHLH--QTLLERKTYHCMPGCKKSFT 249  
 QY 241 KNELVHVNIPHDGNIP---DDLKEVEK-KLENLLDQSKLANLHELETEKLVKVEDE 296  
 DB 250 SSALKKHTSVIHGNWAFHCDSCGKFGYKELQRLHLEGT-CRKAHKPYINEGIRKHDG 308  
 QY 297 EDEDSLDKSDVRSDSMS--AQRSTKSFSTASLEGSKSVKLSNGGKINKPCPNKCDR 354  
 DB 309 IEGVAIHQKELSSNLVSDVAKKIINEVTG--HGKY-----PAREYSCSPFCNY 358  
 QY 355 MFSREYDLRRHLKWH 369  
 DB 359 RFKRLYDVRHLNSH 373  
 RESULT 3  
 ID\_ZN43 HUMAN  
 AC P17038; P28160; Q96DGL; PRT; 803 AA.  
 DT 01-AUG-1990 (Rel. 15; Created)  
 DT 01-DEC-1992 (Rel. 24; Last sequence update)  
 DT 10-OCT-2003 (Rel. 42; Last annotation update)  
 DE Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein K0X27).  
 DE K0X27.  
 GN ZNF43 OR ZNF39 OR K0X27.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-T-cell.  
 RX MEDLINE=91279444; PubMed=1711675;  
 RA Lovering R., Trowsdale J.;  
 RT "A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines."  
 RL Nucleic Acids Res. 19:2921-2927(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;

RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulláhy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 38-190 FROM N.A.  
 RX MEDLINE=91219421; PubMed=2023909;  
 RA Belfroid E.J., Poncelet D.A., Lecocq P.J., Revelant O., Martial J.A.;  
 RT "The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins."  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).  
 RN [4]  
 RP SEQUENCE OF 476-531 FROM N.A.  
 RC TISSUE=Lymphoid;  
 RX MEDLINE=91145339; PubMed=2288909;  
 RA Thiesen H.-J.;  
 RT "Multiple genes encoding zinc finger domains are expressed in human T cells."  
 RL New Biol. 2:363-374(1990).  
 CC -!- FUNCTION: May be involved in transcriptional regulation.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- TISSUE SPECIFICITY: T AND B CELL LINES.  
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.  
 CC -!- SIMILARITY: Contains 1 KRAB domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X59244; CAA41932.1; -  
 CC EMBL; BC006528; AAH0526.1; -  
 CC EMBL; M61869; AAAS674.1; -  
 CC EMBL; X52358; CAA36584.1; -  
 CC PIR; S26823; S26823.  
 CC HSP; P08048; 7ZNF.  
 CC TRANSFAC; T04986; -  
 CC Genew; HGNC:13109; ZNF43.  
 CC MIM; 603972; -  
 CC GO; GO:0003677; F:DNA binding; TAS.  
 CC InterPro; IPR001909; KRAB.  
 CC InterPro; IPR007087; Znf\_C2H2.  
 CC InterPro; IPR007086; Znf\_C2H2\_sub.  
 CC Pfam; PF00096; zf-C2H2; 21.  
 CC PRINTS; PR00048; ZINCFINGER.  
 CC ProDom; PD000003; Znf\_C2H2; 16.  
 CC SMART; SM00349; KRAB; 1.  
 CC SMART; SM00355; Znf\_C2H2; 22.  
 CC PROSITE; PS00805; KRAB; 1.  
 CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 19.



```

Db      321  EKPYKCECGKAFQSSTLTTHKI - IHTGEFKYCEB - CGKAFSQTSLTTHKRIHSGE 377
QY      195  PKLKPCGCGKCGVCKGGLSSHMLSHDSTMLKIWTCDYCDVGKFAKKNELVEHYNI PHDG 254
Db      378  KPYKCECGKAFKQSSSTLTTHKRIHAGE - --KFYKCEVCSCA - PSFSLTTHKRI - HTG 432
QY      255  NIPDLKTEVEVKLENLDDGSKLNLHLETEKLVDEDEDESDLEKRSQVRSDS 314
Db      433  EKP - --YKCECGKAFNLSSTLTTHKIHTGE - KPYKCEB - --GRAFNQSS 478
QY      315  MSAQRSI - --KSFASLEGSKSVKLSNGKKINCPKNNCDRMFRSEYD 361
Db      479  LSKHVIHTGEKPKYKCEGKAFNQS - --SLTTHKMIHTGEKPKYC - --ECGKAFNSSI 534
QY      362  LRRHLKWH - --DNLQRIEFLNSIEKEE 387
Db      535  LNRKHMHTGEKLYKPSNACDNIKISKYKNCAGEK 574

RESULT 5
TF3A_HUMAN
AC      Q92664; Q12963; Q13097; PRT; 423 AA.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Transcription factor IIIA (Factor A) (TFIIIA).
GN      TF3A.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Fetal brain;
RX      MEDLINE=95303028; PubMed=7789179;
RA      Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,
RT      Fujiwara T., Takahashi E., Shin S., Nakamura Y.;
RT      "Molecular cloning, characterization, and chromosomal mapping of a
RT      novel human gene (TF3A) that is highly homologous to Xenopus
RT      transcription factor IIIA.";
RL      Cytogenet. Cell Genet. 70:235-238(1995).
RN      [2]
RP      SEQUENCE OF 61-423 FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=95347600; PubMed=7622052;
RA      Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.,
RA      Becker K.G.;
RT      "Cloning and expression analysis of a human cDNA homologous to
RT      Xenopus TFIIIA.";
RL      Gene 159:215-218(1995).
RN      [3]
RP      CHARACTERIZATION.
RX      MEDLINE=94342241; PubMed=8063702;
RA      Moorefield B., Roeder R.G.;
RT      "Purification and characterization of human transcription factor
RT      IIIA.";
RL      J. Biol. Chem. 269:20857-20865(1994).
CC      -!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC      APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC      CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC      BINDS THE TRANSCRIBED 5S RNA'S. MAY INITIATE TRANSCRIPTION OF THE
CC      5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION
CC      OF OTHER GENES.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- TISSUE SPECIFICITY: Ubiquitous.

```

```

CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D32257; BAA06988.1; -.
DR      EMBL; U20272; AAA75623.1; -.
DR      EMBL; U14134; AAA21873.1; -.
DR      HSSP; P03001; 1TF3.
DR      TRANSFAC; T04953; -.
DR      Genew; HGNC:4662; GTF3A.
DR      MIM; 600860; -.
DR      GO; GO:0003709; F:RNA polymerase III transcription factor act. . .; TAS.
DR      GO; GO:0003030; P:RNA transcription; TAS.
DR      GO; GO:0006383; P:transcription from Pol III promoter; TAS.
DR      InterPro; IPR007087; Znf C2H2.
DR      Pfam; PF00096; zf-C2H2; 9.
DR      SMART; SM00355; Znf C2H2; 9.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR      PROSITE; PS00557; ZINC_FINGER_C2H2_2; 8.
KW      Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW      RNA-binding; Repeat; Nuclear protein; Polymorphism.
FT      ZN_FING 98 122
FT      ZN_FING 128 152
FT      ZN_FING 158 183
FT      ZN_FING 190 212
FT      ZN_FING 220 244
FT      ZN_FING 247 271
FT      ZN_FING 275 297
FT      ZN_FING 304 329
FT      ZN_FING 335 359
FT      VARIANT 303 303
FT      CONFLICT 214 214
FT      CONFLICT 221 221
FT      CONFLICT 223 248
FT      CONFLICT 378 387
SQ      SEQUENCE 423 AA; 46847 MW; A627D064A43FB6F0 CRC64;

Query Match      17.8%; Score 397.5; DB 1; Length 423;
Best Local Similarity 31.1%; Pred. No. 9.9e-17;
Matches 99; Conservative 55; Mismatches 99; Indels 59; Gaps 16;

QY      8  KSISL-----ISSSSSR-----PKYICYECDCAYNRPSLLEOHLRTHSNDR 53
Db      67  ESVSLTTADAFIAAGESSAPTPRPALPRRFGSPDCSANYSKAWKLDALCKHTGER 126
QY      54  PYKCTVDDCDKAFKSHLETHYVSHSEKSPHCSV--CGKGVNSRQHLKREH-EITH--- 107
Db      127  PFVCDYEGCGKAFIRDYHLRHLTHTGKDFVCAANGCDQKFNKTKSLKKHFKKHENQ 186
QY      108  TKSFCKTFCNCEAFYKQSLR-HHLSVHEKTLTKQ--CNKVFTRPSKLACHKLKHG 164
Db      187  OKQYICSFEDCKCTFKHQQLKIHOQNTNEPLFKCTQEGCGKHFASPSKLKRAKAHEG 246
QY      165  GSPAYQCDHPGCFKNTQVSWVLQPHIKOSHPK----- 196
Db      247  ----YVC-QKGCFSVAKTWTLLKHVRETHKEILCEVCRKTKRKYDKYLKQHKMTHAPER 301
QY      197  --LKCPK--CGKGVGKGLSSHMLS-HDDSTMIKWTCDYCDVGK-PAKKNELVEHYNI 250
Db      302  DVCRCPREGCGRTYTVFNLSHLSFHEES---RPFVCEHAGCGKTFAMQSLTRH-AV 357
QY      251  PHDGNIPDLKTEVEVK 268
Db      358  VHD---PDKMKMKLVKK 372

RESULT 6
ZN85_HUMAN
ID      ZN85_HUMAN STANDARD; PRT; 595 AA.
AC      Q03923;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-OCT-1995 (Rel. 34, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)

```

[illegible]

RT "Sequence analysis of a 5.7 Mb region in 19p13.1 between OLFR and  
 RL D9S455.", (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-191 FROM N.A.  
 RX MEDLINE=91219421; PubMed=2023909;  
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,  
 RA Martial J.A.;  
 RT "The evolutionarily conserved Kruppel-associated box domain defines a  
 RT subfamily of eukaryotic multifingered proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).  
 CC -|- FUNCTION: May be involved in transcriptional regulation.  
 CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -|- DEVELOPMENTAL STAGE: Expressed early during embryonic development.  
 CC -|- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -|- SIMILARITY: Contains 1 KRAB domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AC007204; AAD22981.1; .  
 DR EMBL; M61873; AAA83548.1; .  
 DR HSPB; P08046; IALH.  
 DR Genew; HGNC:13169; ZNF93.  
 DR MIM; 603975; .  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0003700; F:transcription factor activity; NAS.  
 DR GO; GO:0008270; F:zinc ion binding; NAS.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
 DR InterPro; IPR001909; KRAB  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR InterPro; IPR007086; Znf\_C2H2\_sub.  
 DR Pfam; PF01352; KRAB; 1.  
 DR Pfam; PF00096; zf-C2H2; 16.  
 DR PRINTS; PR00048; ZINC\_FINGER.  
 DR ProDom; PD000003; Znf\_C2H2; 7.  
 DR SMART; SM00349; KRAB; 1.  
 DR SMART; SM00355; Znf\_C2H2; 16.  
 DR PROSITE; PS00805; KRAB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 16.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 17.  
 KW Transcription regulation; zinc-finger; DNA-binding; Metal-binding;  
 KW Nuclear protein; Repeat.  
 FT NON\_TER 1  
 FT DOMAIN <1 71  
 FT ZN\_FING 141 163  
 FT ZN\_FING 169 191  
 FT ZN\_FING 197 219  
 FT ZN\_FING 225 247  
 FT ZN\_FING 253 275  
 FT ZN\_FING 281 303  
 FT ZN\_FING 309 331  
 FT ZN\_FING 337 359  
 FT ZN\_FING 365 387  
 FT ZN\_FING 393 415  
 FT ZN\_FING 421 443  
 FT ZN\_FING 449 471  
 FT ZN\_FING 477 499  
 FT ZN\_FING 505 527  
 FT ZN\_FING 533 555  
 FT ZN\_FING 561 583  
 FT ZN\_FING 589 611  
 FT ZN\_FING 71 71  
 FT CONFLICT 85 85  
 FT CONFLICT 115 115  
 FT CONFLICT 132 132  
 FT CONFLICT 149 149

SQ SEQUENCE 616 AA; 70572 MW; DF309883AB61160A CRC64;  
 Query Match 17.7%; Score 395; DB 1; Length 616;  
 Best Local Similarity 29.7%; Pred. No. 2.1e-16;  
 Matches 123; Conservative 42; Mismatches 157; Indels 92; Gaps 20;  
 QY 23 KKYICTEGCDKAYNRPSLLRHLRTHNSNDPRPKYCTVDDDCDAFKRSHLETHIVSHSEK 82  
 DB 195 KPYIC--EECGKAFKYSALNTHKRIHTGEKPKC--DKCDKAFIASSTLSKHEIHTGK 250  
 QY 83 KPFHSCVCGKGVNSRQHLKRHEIHT--KSPKCTFENCQEAFFKQSLRHH----- 131  
 DB 251 KPYKECECGKAFNOSSTLTGKHKTHTGEKPKC--EECGKAFNOSSTLTGKHKTHTGKPK 308  
 QY 132 -----ILSVH-----EKLTKQCNKVFTSPSKLAQKHLKH----- 162  
 DB 309 YVCEBEGKAFKYSRLTTHKRIHTGEKPKCNKCGKAFIASSTLSRHEFIHMGKHKKE 368  
 QY 163 -----HGSPAYQCDHPGCPKFNFTWSVLQFHKOSH--PKLKCPK 201  
 DB 369 ECGKAFIWSVLTTRHKRYHTGEKPKCEE--CGKAFKYSSTLSH--KRSHTEGKPKCEE 425  
 QY 202 CGKGVGKGLSSHMLSHDDSTMIKIWTCDYCDVGK-PAKKNELVEHYNIHFDGNIPDDL 260  
 DB 426 CGKAFVASTLSKHEIHT--TGKPKYKCEC--GKAFNQSSSLTKKKI-HTGKPK-- 476  
 QY 261 LKTEVKKLENLLDQSKLNNLHELET--EKLKVEDEDEDEDS---LDEKRSVRSDSM 315  
 DB 477 YKCECGKAFN---QSSSLTKHKHTGEKPKYKCECGKAFNQSSSLTKKKIHTREKPY 533  
 QY 316 SAQRSIKSFTASLEGSKSVKLSNGKKINCPKNCMDRPFMSREYDLRRHLKWH 369  
 DB 534 KCECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSHKKI 583  
 RESULT 8  
 ZN91 HUMAN  
 ID ZN91 HUMAN STANDARD; PRT; 1191 AA.  
 AC Q05481;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 91 (Zinc finger protein HTF10) (HPF7).  
 GN ZNF91.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93223677; PubMed=8467795;  
 RA Bellefroid E.J., Marine J.C., Ried T., Lecocq P.J., Riviere M.,  
 RA Anemiy C.T., Poncelet D.A., Coullie P.G., de Jong P.J.,  
 RA Szpirer C., Ward D.C., Martial J.A.;  
 RT "Clustered organization of homologous KRAB zinc-finger genes with  
 RT enhanced expression in human T lymphoid cells.";  
 RL EMBO J. 12:1363-1374(1993).  
 RN [2]  
 RP SEQUENCE OF 15-204 FROM N.A.  
 RX MEDLINE=91219421; PubMed=2023909;  
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,  
 RA Martial J.A.;  
 RT "The evolutionarily conserved Kruppel-associated box domain defines a  
 RT subfamily of eukaryotic multifingered proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).  
 CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -|- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -|- SIMILARITY: Contains 1 KRAB domain.  
 CC -|- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS  
 CC DERIVED FROM THE TRANSLATION OF AN ALU REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----

DR EMBL; J11672; AAA59469.1; -;  
DR EMBL; M61871; AAA58672.1; ALT\_SEQ.  
DR PIR; S35305; S35305.  
DR HSSP; P08047; LSP2.  
DR Genew; HGNC:13166; ZNF91.  
DR MIM; 603971; -;  
DR CO; GO:0005634; C-nucleus; NAS.  
DR CO; GO:0003700; Transcription factor activity; NAS.  
DR GO; GO:0008270; Zinc ion binding; NAS.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
DR InterPro; IPR001909; KRAB.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF01352; KRAB; 1.  
DR Pfam; PF00096; zf-C2H2; 34.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR ProDom; PD000003; Znf\_C2H2; 20.  
DR SMART; SM00349; KRAB; 1.  
DR SMART; SM00355; Znf\_C2H2; 33.  
DR PROSITE; PS00805; KRAB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 31.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 35.  
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;  
KW Nuclear protein; Repeat.  
FT DOMAIN 13 84  
FT ZN\_FING 154 176 C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 179 200 C2H2-TYPE.  
FT ZN\_FING 208 232 C2H2-TYPE.  
FT ZN\_FING 238 260 C2H2-TYPE.  
FT ZN\_FING 266 288 C2H2-TYPE.  
FT ZN\_FING 294 316 C2H2-TYPE.  
FT ZN\_FING 322 344 C2H2-TYPE.  
FT ZN\_FING 350 372 C2H2-TYPE.  
FT ZN\_FING 378 400 C2H2-TYPE.  
FT ZN\_FING 406 428 C2H2-TYPE.  
FT ZN\_FING 434 456 C2H2-TYPE.  
FT ZN\_FING 462 484 C2H2-TYPE.  
FT ZN\_FING 490 512 C2H2-TYPE.  
FT ZN\_FING 518 540 C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 546 568 C2H2-TYPE.  
FT ZN\_FING 574 596 C2H2-TYPE.  
FT ZN\_FING 602 624 C2H2-TYPE.  
FT ZN\_FING 630 652 C2H2-TYPE.  
FT ZN\_FING 658 680 C2H2-TYPE.  
FT ZN\_FING 686 708 C2H2-TYPE.  
FT ZN\_FING 714 736 C2H2-TYPE.  
FT ZN\_FING 742 764 C2H2-TYPE.  
FT ZN\_FING 770 792 C2H2-TYPE.  
FT ZN\_FING 798 820 C2H2-TYPE.  
FT ZN\_FING 826 848 C2H2-TYPE.  
FT ZN\_FING 854 876 C2H2-TYPE.  
FT ZN\_FING 885 904 C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 910 932 C2H2-TYPE.  
FT ZN\_FING 938 960 C2H2-TYPE.  
FT ZN\_FING 966 988 C2H2-TYPE.  
FT ZN\_FING 994 1016 C2H2-TYPE.  
FT ZN\_FING 1022 1044 C2H2-TYPE.  
FT ZN\_FING 1050 1072 C2H2-TYPE.  
FT ZN\_FING 1078 1100 C2H2-TYPE.  
FT ZN\_FING 1106 1128 C2H2-TYPE.  
FT ZN\_FING 1134 1156 C2H2-TYPE.  
SQ SEQUENCE 1191 AA; 137225 MW; 581056BBI1B8716D CRC64;

Query Match 17.7%; Score 393.5; DB 1; Length 1191;  
Best Local Similarity 35.8%; Pred. No. 5.2e-16;  
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGCDKAYNPSLLEQHLRTHSNDRPYKCTVDDCKAFKPKSHLETHIVSH 79  
DB 905 TREKPYKC--EECGKAPSQFSLTTHKRWHTGEKPYKC--EECGKAPFSQSSTLTTHKIIH 960  
QY 80 SEKKPFHSCVCGKGVNSRQHLKREHETHT--KSFKCTFENCQEAFFYKHQSL-RHHILSVH 136  
DB 961 TGEKPYKCECGKAPKPSSTLTTHKIIHTGEKPYKC--EECGKAPFSQSSTLTTHKTHMTG 1018  
QY 137 EKTLTKQCNKVFTRPSKLAHKLKHGGSPAYQCQDHPGCFKNFQWLSVLPQHIKQSHPK 196  
DB 1019 EKPYKCECGKAPKPSSTLTTHKIIHTGEKPYKC--EECGKAPFSQSSTLTTHKTHMTG 1074  
QY 197 ---LKCPKCGKCGVGGKGLSSHMLSHDDSTMIKWTCDYCDVGKFAKKNELVEHYNIHFD 253  
DB 1075 EKPYKCECGKAPFSQSSTLTTHKRLH---TGEKPYKCGEC--GKAFKSSALTTHKIIHT 1129  
QY 254 GNPDDLKTEVKKLENLLDQSKLNNLHLETF 287  
DB 1130 GE-----KPYKCEKCKCAFQSSILTNHKKIHT 1157  
RESULT 9  
Z347\_HUMAN STANDARD; PRT; 839 AA.  
ID 2347\_HUMAN STANDARD; PRT; 839 AA.  
AC Q96SE7; Q8TON1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc finger protein 347 (Zinc finger 1111).  
GN ZNF347 OR ZNF1111.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RA Aiken C.J., Nicholson G.C.;  
RT "RANKL regulated zinc finger protein in osteoclastogenesis.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE OF 305-839 FROM N.A.  
RC TISSUE=Brain;  
RA Bloeker H., Boscher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: May function as a transcription factor.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.  
CC -!- SIMILARITY: Contains 1 KRAB domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; AY029765; AAK37403.1; -;  
EMBL; AL713691; CAD28491.1; -;  
DR Genew; HGNC:16447; ZNF347.  
DR InterPro; IPR001909; KRAB.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF01352; KRAB; 1.  
DR Pfam; PF00096; zf-C2H2; 18.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR ProDom; PD000003; Znf\_C2H2; 2.  
DR SMART; SM00349; KRAB; 1.  
DR SMART; SM00355; Znf\_C2H2; 19.  
DR PROSITE; PS00805; KRAB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 17.

DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 20.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
 KW Nuclear protein; Repeat.

FT DOMAIN 8 79

FT ZN\_FING 261 283 C2H2-TYPE (DEGENERATE).

FT ZN\_FING 289 311 C2H2-TYPE.

FT ZN\_FING 317 339 C2H2-TYPE.

FT ZN\_FING 345 367 C2H2-TYPE.

FT ZN\_FING 373 395 C2H2-TYPE.

FT ZN\_FING 401 423 C2H2-TYPE.

FT ZN\_FING 429 451 C2H2-TYPE.

FT ZN\_FING 457 479 C2H2-TYPE.

FT ZN\_FING 485 507 C2H2-TYPE.

FT ZN\_FING 513 535 C2H2-TYPE.

FT ZN\_FING 541 563 C2H2-TYPE.

FT ZN\_FING 569 591 C2H2-TYPE.

FT ZN\_FING 597 619 C2H2-TYPE.

FT ZN\_FING 625 647 C2H2-TYPE (DEGENERATE).

FT ZN\_FING 653 675 C2H2-TYPE.

FT ZN\_FING 681 703 C2H2-TYPE.

FT ZN\_FING 709 731 C2H2-TYPE.

FT ZN\_FING 737 759 C2H2-TYPE.

FT ZN\_FING 765 787 C2H2-TYPE.

SQ SEQUENCE 839 AA; 95771 MW; F77C7069E7844E CRC64;

Query Match 17.4%; Score 387.5; DB 1; Length 839;

Best Local Similarity 30.7%; Pred. No. 8e-16;

Matches 112; Conservative 46; Mismatches 154; Indels 53; Gaps 16;

11 SSL-SSSSSSPKKYICYEGCDKAYNRPSSLECHLRTHNDPRYKTVDDCDKAFERK 69

386 SSLAIHQATHSGEKPKYC--NECGKVFQNSHLTNHRLHGTGEKPKC--NECGKAFGVR 441

70 SHLETHIVSHSEKKYFHCVSCKGKNSRQHLKRHEITHT--KSPKCTFENCQBAFYKHQS 127

442 SSLAIHLVHTGEKPKYKCHCKGKVFRRNSHLARHQLHTGKPKYC--NECGKAFRAHSN 499

128 L-RHILSVHEHTLTCQCNKVFTRPSKLAQHLKHGGSPAYQCDHPGCFKFNQVSWVL 186

500 LITHQVIHTGEKPKYKNSCGKVFQNSHLANQRIHTGVKP-YCNS--CGKAFSVYSSL 556

187 QFH--IKSHPKLKPCKGKGVCKGKGLSSHLSDOSTMIKIWTQCYDVGKFAKNEL 244

557 TTHQVIHTGEKPKYKNSCGKVFQNSHLARHGIH--TGKPKYKNEC--GKVFRENSY 611

245 VHYNIPIFDGNIPIDDLKETEVEKLENLLDQSKLNNLHELETKLVEDEDEDESDLD 304

612 LSRHQELHTGKXP---YKNEYGK---AFSEHSLTTHQVIHTGEKPKYKNE----- 657

305 EKRSVRSDSMAQSIKSFSTASLEGSKSVKLSNSGKKINCPCNNCDRMFGREYDLR 364

658 -----CGKVFQNSHLAR--HRRVHTGKPKYC--NECGKAFSGTSKLAR 698

365 HLKWH 369

699 HORVH 703

RESULT 10

2431 HUMAN

ID Z431 HUMAN STANDARD; PRT; 576 AA.

AC Q8TF32; Q8TWC4;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DE Zinc finger protein 431.

GN ZNF431 OR KIAA1969.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1] SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA MEDLINE=21842142; PubMed=11853319;

RA Nagase T., Kikuno R., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes, XXII.

The complete sequences of 50 new cDNA clones which code for large

proteins.";

RNA Res. 8:319-327 (2001).

[2]

SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA MEDLINE=22388257; PubMed=12477932;

RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton S., Kettner M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).

CC -!- FUNCTION: May be involved in transcriptional regulation.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-

FINGER PROTEINS.

CC -!- SIMILARITY: Contains 1 KRAB domain.

CC -!- SIMILARITY: Contains 13 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)

or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL; AB075849; BAB85555.1; ALT\_INIT.

CC EMBL; BC040506; AAH40506.1; -.

CC Genew; HGNC:20809; ZNF431.

CC InterPro; IPR001909; KRAB.

CC InterPro; IPR007087; Znf\_C2H2.

CC InterPro; IPR007086; Znf\_C2H2\_sub.

CC Pfam; PF01352; KRAB; 1.

CC PRINTS; PR00048; ZNCFINGER.

CC SMART; SM00349; KRAB; 1.

CC SMART; SM00355; Znf\_C2H2; 12.

CC PROSITE; PS00805; KRAB; 1.

CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 12.

CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 13.

CC Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;

CC Nuclear protein; Repeat.

CC DOMAIN 35 106

CC ZN\_FING 176 198

CC ZN\_FING 204 226

CC ZN\_FING 232 254

CC ZN\_FING 260 282

CC ZN\_FING 288 310

CC ZN\_FING 316 338

CC ZN\_FING 344 366

CC ZN\_FING 372 394

CC ZN\_FING 400 422

CC ZN\_FING 428 450

CC KRAB

CC C2H2-TYPE 1 (DEGENERATE).

CC C2H2-TYPE 2.

CC C2H2-TYPE 3.

CC C2H2-TYPE 4.

CC C2H2-TYPE 5.

CC C2H2-TYPE 6.

CC C2H2-TYPE 7.

CC C2H2-TYPE 8.

CC C2H2-TYPE 9.

CC C2H2-TYPE 10.

```
FT ZN_FING 456 478 C2H2-TYPE 11.
FT ZN_FING 484 506 C2H2-TYPE 12.
FT ZN_FING 512 534 C2H2-TYPE 13.
FT CONFLICT 198 198 H -> R (IN REF. 2).
SQ SEQUENCE 576 AA; 67216 MW; 532774BF69EC9E2A CRC64;

Query Match 17.3%; Score 385; DB 1; Length 576;
Best Local Similarity 29.4%; Pred. No. 7,5e-16;
Matches 109; Conservative 48; Mismatches 156; Indels 56; Gaps 15;

QY 8 KSISSLISSSSSS---RPKKYICTYEGCDKAYNRPSSLLEQHLRTHSNDRPYKCTVDDCD 63
Db 211 KSPFCLLHLSQKRIHIRENSYQC--EECGKAFKFWSTLTKRIHTGKPKFC--EECG 266
QY 64 KAPFKSHLETHIVSHSEKPPHCVGKGVNSROHLKREHITHT--KSPKCTFENCQEA 121
Db 267 KAFKOSSTLTTHKRIHTGKPKFC--EECGKAFNRSSHLTTHKRIHTGKPKFC--EECGKA 324
QY 122 FYKHOSL-RHHILSVHEKTLTQCQNKVTPRPSKLAQHLKHGGSPAYQCDHPGCFKX 180
Db 325 FNGSSTLSTHKTIFHAGEKPKYKCECDKAFNRSYLTKKI-IHTGKSYKCE--CGKGF 381
QY 181 QTVSWVLQFH--IKQSHPKLKCPKCGKGVGKGLSSHMLSHDDSTMKIWTCDYCDVCKP 238
Db 382 NWSSTLTGKRIHTGKPKYKCEGKAFNRSNLTTHKMIH---TGKPKYKCEC--GKA 436
QY 239 AKKNELVEHYNIPHDGNI PDDLKXETEVKLEMLDQSGKLNHLHELETKLVDEED 298
Db 437 FNRSPQLTAHKIHTGE-----KPKYKCEGKAFSQSSILTT----- 473
QY 299 EEDSLDEKSDVRSQMSAQRIKFTASLEGSKSVLSKINSKINKCPKNCNDRMPSR 358
Db 474 -----HKRIHTGKPKYKCEGKAFNRSNLTTHKMIH---HKIHTGKSYK--EECGKAFNQ 523
QY 359 EYDLRHLKWH 369
Db 524 SSTLKHKRIH 534

RESULT 11
ID 2228 HUMAN STANDARD; PRT; 913 AA.
AC Q9UJ3; O9HCA7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 228.
GN ZNF228.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Strubbs L.;
RT "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Kodoyanni V., Ge Y., Krummel G.K., Kvietad E., Grable L.,
RA Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
RT gene cluster.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 1 KRAB domain.
CC -!- SIMILARITY: Contains 17 C2H2-type zinc fingers.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF198358; AAF12816.1; -
CC EMBL; AC084239; AAG23968.1; -
CC HSP; P08047; ISP2.
CC Genew; HGNC:13021; ZNF228.
CC InterPro; IPR001909; KRAB.
CC InterPro; IPR007087; Znf_C2H2.
CC InterPro; IPR007086; Znf_C2H2_sub.
CC Pfam; PF00036; zf-C2H2; 13.
CC PRINTS; PR00048; ZINC_FINGER.
CC PRODOM; PD000003; Znf_C2H2; 12.
CC SMART; SM00349; KRAB; 1.
CC SMART; SM00355; Znf_C2H2; 13.
CC PROSITE; PS00805; KRAB; 1.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 17.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein.
FT DOMAIN 8 79 KRAB.
FT ZN_FING 258 280 C2H2-TYPE 1 (DEGENERATE).
FT ZN_FING 443 465 C2H2-TYPE 2 (DEGENERATE).
FT ZN_FING 471 493 C2H2-TYPE 3 (DEGENERATE).
FT ZN_FING 497 519 C2H2-TYPE 4 (DEGENERATE).
FT ZN_FING 525 547 C2H2-TYPE 5.
FT ZN_FING 553 575 C2H2-TYPE 6.
FT ZN_FING 581 603 C2H2-TYPE 7.
FT ZN_FING 609 631 C2H2-TYPE 8.
FT ZN_FING 637 659 C2H2-TYPE 9.
FT ZN_FING 665 687 C2H2-TYPE 10.
FT ZN_FING 693 715 C2H2-TYPE 11.
FT ZN_FING 721 743 C2H2-TYPE 12.
FT ZN_FING 749 771 C2H2-TYPE 13.
FT ZN_FING 777 799 C2H2-TYPE 14.
FT ZN_FING 805 827 C2H2-TYPE 15.
FT ZN_FING 833 855 C2H2-TYPE 16.
FT ZN_FING 861 883 C2H2-TYPE 17.
FT CONFLICT 3 VSK -> KFO (IN REF. 2).
FT CONFLICT 113 T -> A (IN REF. 2).
FT CONFLICT 143 A -> V (IN REF. 2).
FT CONFLICT 222 E -> K (IN REF. 2).
FT CONFLICT 261 S -> T (IN REF. 2).
FT CONFLICT 392 S -> N (IN REF. 2).
FT CONFLICT 446 Q -> E (IN REF. 2).
FT CONFLICT 485 H -> Y (IN REF. 2).
FT CONFLICT 813 A -> G (IN REF. 2).
SQ SEQUENCE 913 AA; 105702 MW; 8D9F5D27CD9FA6B CRC64;

Query Match 17.2%; Score 383; DB 1; Length 913;
Best Local Similarity 30.7%; Pred. No. 1.6e-15;
Matches 109; Conservative 55; Mismatches 133; Indels 58; Gaps 19;

QY 23 KKYICTYEGCDKAYNRPSSLLEQHLRTHSNDRPYKCTVDDCDKAFKSHLETHIVSHSEK 82
Db 551 KPYKFC--EECDKGFSSSYLAQHQVHTGKPKFC--EECGKGFSSSYLAQHQVHTG 606
QY 83 KPFHCSYCGKGVNSROHLKREHITHT--KSPKCTFENCQEAIFYKHOSL-RHHILSVHEKT 139
Db 607 KPYKCECGKGFSSSYLAQHQVHTGKPKFC--EECGKGFSSSYLAQHQVHTGKPK 664
QY 140 LTCQCKNVTPRPSKLAQHLKHGGSPAYQCDHPGCFKXQTVSWVLQFH--IKQSHPKL 197
Db 665 YKCECGKGFSSSYLAQHQVHTGKPKFC--EECGKGFSSSYLAQHQVHTGKPK 721
QY 198 KCPKCGKGVGKGLSSHMLSHDDSTMKIWTCDYCDVKGKFAKNELVEHYNIPHDGNI 257
```

```

Db 722 ICEVCGKFSQRAYLQGHQVH---TRVKPKCEMCKGK-PSOSSRLEAHRV-HTGGKP 776
QY 258 DLLKETEYKLENNLLDOGSKLNNHLETKLKVVEDEDEDSLEKSDVRSDSMSA 317
Db 777 -----YKCE-VCTKGF-----ESSRLQAHQ-----RVHVEGRPYKC 807
QY 318 QRSIKSFT--ASLEGSVSKLISNGKK-INCPKNNCDRMFSREYDLRLHLKWH 369
Db 808 EQCKAFSGYSLQAHHRV-----HTGEKPYKC--EVCCKGFSORSNLQAHQVH 855

RESULT 12
ID 2430 HUMAN STANDARD; PRT; 570 AA.
AC Q9H8G1; OS6V70;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Zinc finger protein 430.
GN ZNF430.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RA Bi A., Yu L.;
RT "Homo sapiens mRNA similar to zinc finger protein 85.";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN 2
RP SEQUENCE OF 1-364 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
RA Nagatsuma M., Haseoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: May be involved in transcriptional regulation.
CC 1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC 1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC 1- SIMILARITY: Contains 1 KRAB domain.
CC 1- SIMILARITY: Contains 12 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY269787; AAP30885.1; -
DR EMBL; AK023721; BAB14656.1; -
DR Genew; HGNC:20808; ZNF430.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf_C2H2; 3.
DR ProDom; PD000003; Znf_C2H2; 3.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
FT DOMAIN 35 107
FT ZN_FING 205 227

```

```

FT ZN_FING 233 255
FT ZN_FING 261 283
FT ZN_FING 289 311
FT ZN_FING 317 339
FT ZN_FING 345 367
FT ZN_FING 373 395
FT ZN_FING 401 423
FT ZN_FING 429 451
FT ZN_FING 457 479
FT ZN_FING 485 507
FT ZN_FING 513 535
FT CONFLICT 293 364
SQ SEQUENCE 570 AA; 66346 MW; 0561E871DC92B61F CRC64;
Query Match 16.9%; Score 376; DB 1; Length 570;
Best Local Similarity 32.5%; Pred. No. 2.5e-15;
Matches 94; Conservative 38; Mismatches 109; Indels 48; Gaps 13;
QY 23 KKVICYEGCDKAYNRPSLLEQHLRTHSNDREPKYCTVDDCDKAFPRKSHLETHIVSHSEK 82
Db 287 KYRC--EECGKTFNRSSHLTHKRIHTEGKEPYRC--EECGAFNRSSHLTHKRIHTEG 342
QY 83 KPFCVCGKGVNSROHLKRHEITHT--KSPFCTFENCQEAIFYKQSL-RHILSVHEK- 138
Db 343 KPYKCECGKAFNQSSSTLTTHKIIHAGEKPYKC--EECGAFYRFSYLTTHKIIHTEGKF 400
QY 139 -----TLT-----CKQCNKVFTRPSKLAQHKLKHGGSPAYOC 171
Db 401 KYCECGKGFNWSSTLTTHKRIHTEGKEPYKCQCGKAFNENSLTAHKIIHTEGKEP-YKC 459
QY 172 DHPGCFKFNQTVSVLQFH--IKQSHPKLCKPKGKGVCKGLSSHLSHSDSTMIKIWT 229
Db 460 EE--CGKAFNRSPKLTAKHVIHSGEKPYPKCECGKAFNQFSLTKHKIHTGDTSYKYLE 517
QY 230 CDYCDVGKAKKNELVEHVNIEDGNIPDLKETEY-KLENLLDDQS 277
Db 518 CDKA----FSQSSTLTGH-KVIHTEGKPYNCEYKAFNQSSNLIZQSN 561

RESULT 13
Z184 HUMAN STANDARD; PRT; 751 AA.
ID Z184 HUMAN
AC Q99676; OS60792;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 184.
GN ZNF184.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN 2
RP SEQUENCE OF 26-751 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97230463; PubMed=9073517;
RA Goldwurm S., Menzies M.L., Banyer J.L., Powell B.L.W.,
RA Jazwinska E.C.;
RT "Identification of a novel Krueppel-related zinc finger gene (ZNF184)
RT mapping to 6p21.3.";
RL Genomics 40:486-489(1997).
CC 1- FUNCTION: May be involved in transcriptional regulation.
CC 1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC 1- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS.
CC 1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.

```

CC CC -!- SIMILARITY: Contains 1 KRAB domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AL021918; CAA17278.1; --  
CC EMBL: U66561; AAC51180.1; --  
CC HSP: P08046; IAH.  
CC Genew: HGNC:12975; ZNF184.  
CC MIM: 602277; --  
CC GO: GO:0005634; C:nucleus; NAS.  
CC GO: GO:0003677; F:DNA binding; NAS.  
CC GO: GO:0003270; F:zinc ion binding; NAS.  
CC GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
CC InterPro: IPR001909; KRAB.  
CC InterPro: IPR007087; Znf\_C2H2.  
CC Pfam: PF01352; zf-C2H2; 19.  
CC PRINTS: PR00048; ZINC\_FINGER.  
CC ProDom: PD000003; Znf\_C2H2; 18.  
CC SMART: SM00349; KRAB; 1.  
CC SMART: SM00355; Znf\_C2H2; 19.  
CC PROSITE: PS00805; KRAB; 1.  
CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 19.  
CC PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 19.  
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
KW Nuclear protein; Repeat.  
FT DOMAIN 28 99 KRAB.  
FT ZN\_FING 222 244 C2H2-TYPE.  
FT ZN\_FING 250 272 C2H2-TYPE.  
FT ZN\_FING 278 300 C2H2-TYPE.  
FT ZN\_FING 306 328 C2H2-TYPE.  
FT ZN\_FING 334 356 C2H2-TYPE.  
FT ZN\_FING 362 384 C2H2-TYPE.  
FT ZN\_FING 390 412 C2H2-TYPE.  
FT ZN\_FING 418 440 C2H2-TYPE.  
FT ZN\_FING 446 468 C2H2-TYPE.  
FT ZN\_FING 474 496 C2H2-TYPE.  
FT ZN\_FING 502 524 C2H2-TYPE.  
FT ZN\_FING 530 552 C2H2-TYPE.  
FT ZN\_FING 558 580 C2H2-TYPE.  
FT ZN\_FING 586 608 C2H2-TYPE.  
FT ZN\_FING 614 636 C2H2-TYPE.  
FT ZN\_FING 642 664 C2H2-TYPE.  
FT ZN\_FING 670 692 C2H2-TYPE.  
FT ZN\_FING 698 720 C2H2-TYPE.  
FT ZN\_FING 726 748 C2H2-TYPE.  
FT CONFLICT 27 27 A -> S (IN REF. 2).  
FT CONFLICT 295 295 Q -> R (IN REF. 2).  
FT CONFLICT 359 359 E -> G (IN REF. 2).  
FT CONFLICT 379 381 QHQ -> PHP (IN REF. 2).  
FT CONFLICT 498 498 R -> G (IN REF. 2).  
FT CONFLICT 526 526 Q -> G (IN REF. 2).  
FT CONFLICT 533 533 K -> E (IN REF. 2).  
FT CONFLICT 547 549 KHE -> QHQ (IN REF. 2).  
FT CONFLICT 746 746 R -> K (IN REF. 2).  
SQ SEQUENCE 751 AA; 86174 MW; 2490FB302C6BC97D CRC64;  
Query Match 16.9%; Score 376; DB 1; Length 751;  
Best Local Similarity 25.2%; Pred. No. 3.4e-15;  
Matches 114; Conservative 52; Mismatches 146; Indels 140; Gaps 20;  
QY 23 KYICTVGGDKAYNRPSLLEQHLRTHSNDRPKCTVDDCDKAFKFKSHLETHIVSHSEK 82  
DB 276 KPYKVC--DQCGKGPIEGFSLTQRIHTGKPKYC--DECGKAFSQRTHLVQRIHTGE 331  
QY 83 KPFHCSVCGKGVNSRQLKRHEITHT--KSFKCTFENCQAFYK-----HQS----- 127

DB 332 KPYTCNECGKAFSGRHFMEHQIHTGKPKYC--DECDKTFRSTHLTHQHKIHTGKKT 389  
QY 128 -----LRHEILSVHEKTLTCKQNKVFTFSPSKLAQHLKHHGGSPAYOC 171  
DB 390 YKNECGKAFNGSPSTIRHMIHTGKPECECGKAFSQHSNLTQHKIHTGKPK-YDC 448  
QY 172 DHFGCPKNTQVTLVQFHK-----QSHPKL-----KCPK 201  
DB 449 AE--CGKSFYSWLSLAQHLKIHTEGKPKYKNECGKAFSYCSSLTQRRHITRREKPECSE 506  
QY 202 CGKGVGKGLSSHMLSHDDSTWIKITWCDYGVGK-FAKNELVEHYNIHFHDGNIP--- 257  
DB 507 CGKAFSVLSNLNQHOKTH---TQEKAYEKEC--GKAFIRSSSLAKHERI-HTGKPYQC 560  
QY 258 -----DDLKETEYVKLE---NLLDQGSKLN-NLHELETEK----- 290  
DB 561 HECGKTFYSSSLTQHKIHTGERPKYKNECGKAFNQNIHLTQHKIHTGAKPYECAECG 620  
QY 291 -----KVEDEDEEDSLDEKSDVRSDSMSAQRSIKSFSTASLEGSKSVSKL 337  
DB 621 KAFHCSLSLAHQHOKTHTEEPYQCNCKEFTFSOSHLTQHQR----- 662  
QY 338 ISNSGKINCKPKNCDRMFSREYDLRRHLKWH 369  
DB 663 IHTGKPKYC--NECDKAFSRSTHLTEHQNTH 692  
RESULT 14  
XFIN XENLA STANDARD; PRT; 1350 AA.  
AC P08045;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc finger protein Xfin.  
GN XENOPUS laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_taxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=88082679; PubMed=2826129;  
RA Ruiz i Altaba A., Perry-O'Keefe H., Melton D.A.;  
RT "Xfin: an embryonic gene encoding a multifingered protein in  
RL Xenopus";  
RL EMBL J. 6:3065-3070 (1987).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=94021366; PubMed=7692399;  
RA Andreazoli M., de Lucchini S., Costa M., Barsacchi G.;  
RT "RNA binding properties and evolutionary conservation of the Xenopus  
RL multifinger protein Xfin";  
RL Nucleic Acids Res. 21:4218-4225 (1993).  
RN [3]  
RP STRUCTURE BY NMR OF FINGER 31.  
RX MEDLINE=89346749; PubMed=2503871;  
RA Lee M.S., Gippert G.P., Soman K.V., Case D.A., Wright P.E.;  
RT "Three-dimensional solution structure of a single zinc finger DNA-  
RL binding domain";  
RL Science 245:635-637 (1989).  
RN [4]  
RP STRUCTURE BY NMR OF A FINGER.  
RX MEDLINE=89378224; PubMed=2506074;  
RA Lee M.S., Cavanagh J., Wright P.E.;  
RT "Complete assignment of the 1H NMR spectrum of a synthetic zinc  
RL finger from Xfin. Sequential resonance assignments and secondary  
RL structure";  
RL FEBS Lett. 254:159-164 (1989).  
CC -!- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL

CC REGULATION PROCESSES.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH  
CC AS NEURAL RETINA CONES.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND  
CC EMBRYOGENESIS.  
CC -!- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.  
CC -!- PTM: Phosphorylated.  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: Contains 1 KRAB domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X04021; CAA29425.1; -.  
CC PDB: 1ZNF; 15-OCT-91. -.  
CC InterPro: IPR001909; KRAB.  
CC InterPro: IPR007087; Znf\_C2H2.  
CC Pfam: PF01352; KRAB; 1.  
CC Pfam: PF00096; zf-C2H2; 36.  
CC ProDom: PD000003; Znf\_C2H2; 20.  
CC SMART: SM00349; KRAB; 1.  
CC SMART: SM00355; Znf\_C2H2; 35.  
CC PROSITE: PS00805; KRAB; 1.  
CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 35.  
CC PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 37.  
CC Zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;  
CC Phosphorylation.  
CC -----  
CC DOMAIN 1 58 KRAB  
FT ZN\_FING 108 130 C2H2-TYPE.  
FT ZN\_FING 136 158 C2H2-TYPE.  
FT ZN\_FING 164 186 C2H2-TYPE.  
FT ZN\_FING 192 214 C2H2-TYPE.  
FT ZN\_FING 220 242 C2H2-TYPE.  
FT ZN\_FING 248 270 C2H2-TYPE.  
FT ZN\_FING 276 298 C2H2-TYPE.  
FT ZN\_FING 326 348 C2H2-TYPE.  
FT ZN\_FING 354 376 C2H2-TYPE.  
FT ZN\_FING 382 404 C2H2-TYPE.  
FT ZN\_FING 410 432 C2H2-TYPE.  
FT ZN\_FING 438 460 C2H2-TYPE.  
FT ZN\_FING 466 488 C2H2-TYPE.  
FT ZN\_FING 503 525 C2H2-TYPE.  
FT ZN\_FING 531 553 C2H2-TYPE.  
FT ZN\_FING 559 581 C2H2-TYPE.  
FT ZN\_FING 587 609 C2H2-TYPE.  
FT ZN\_FING 615 637 C2H2-TYPE.  
FT ZN\_FING 643 665 C2H2-TYPE.  
FT ZN\_FING 671 693 C2H2-TYPE.  
FT ZN\_FING 699 721 C2H2-TYPE.  
FT ZN\_FING 750 772 C2H2-TYPE.  
FT ZN\_FING 778 800 C2H2-TYPE.  
FT ZN\_FING 806 828 C2H2-TYPE.  
FT ZN\_FING 834 856 C2H2-TYPE.  
FT ZN\_FING 862 884 C2H2-TYPE.  
FT ZN\_FING 890 912 C2H2-TYPE.  
FT ZN\_FING 918 940 C2H2-TYPE.  
FT ZN\_FING 988 1010 C2H2-TYPE.  
FT ZN\_FING 1016 1038 C2H2-TYPE.  
FT ZN\_FING 1044 1066 C2H2-TYPE.  
FT ZN\_FING 1136 1158 C2H2-TYPE.  
FT ZN\_FING 1164 1186 C2H2-TYPE.  
FT ZN\_FING 1192 1214 C2H2-TYPE.  
FT ZN\_FING 1220 1242 C2H2-TYPE.  
FT ZN\_FING 1248 1270 C2H2-TYPE.  
FT ZN\_FING 1276 1298 C2H2-TYPE.  
FT STRAND 1045

FT STRAND 1052 1052  
FT HELIX 1056 1062  
FT HELIX 1063 1065  
FT TURN 1066 1066  
SQ SEQUENCE 1350 AA; 155804 MW; 27F10AB08051E0AD8 CRC64;  
  
Query Match 16.8%; Score 375.5; DB 1; Length 1350;  
Best Local Similarity 27.3%; Pred. No. 6.8e-15;  
Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;  
  
QY 11 SSLISSSSRPKYICYEGCDKAYNPSPLEQHLRTHSDRVPYKCTVDDCDKAFERKS 70  
DB 340 SFANHQQTHSREKPYLCSH--CNKGFIQNSDLVKHFRTHTGERPYQCA--ECHKGFQKS 395  
QY 71 HLETHIVSHSEKKPPHCVGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFFYKHOSL 128  
DB 396 DLVXHLRTHTGEKPKKCHCDKKTTERSALAKHQHTHTGEKPKCS--DCGKEFTQSGNL 453  
QY 129 -RHHILSVHEKTLTCQCNKVFTSPSKLAQHLKH-----HGGSPAYQCDHPGCEKN 179  
DB 454 ILHQRIHTGERPKCTCLDRTFIQNSDLVKHQRKVNANPLSDPHTANSFKCSK--CDLT 511  
QY 180 FQTWSVLQPHIK--QSHPKLCKPKCGKGVKGLSSHMLSH-----DD 221  
DB 512 FSHWTFMKHKLHSGEKKFQCAECKKQFTQKSDLVKHIRVHTGTEKPKFKLLCKKFSQN 571  
QY 222 STMKIW-----TCDYCDVGVKFAKNELVEHYNIHFGNIPDDLLKETEYVKLENL 272  
DB 572 SDLHKHRIHTGEKPPCYCTCD-KSFTERSALIKHRT--HTGERPH-----KCSVCQKG 623  
QY 273 LDGSKLNNLHELETEKLKVEDEDEDESDLESKSDVRSMSAQR-----S 320  
DB 624 FTKQKAL-----TMSRTHTGERPKYPCYCTCGKSFQNSDLVKHQRHTGTEKPKYHCTEC 676  
QY 321 IKSFASLEGSKSVSKLINSQKK-INCPKNMCDRMFSREYDLRRLKWHQD-DNIQRLES 378  
DB 677 NK3FT---EGSLVKHRRTHSGEKPKYRCPQ--CEKTFIQSSDLVKHVVHNGENPAAATA 731  
QY 379 FLN-SIEKEETPEGE 393  
DB 732 FHEILIRRENLRSEP 747  
  
RESULT 15  
Z257\_HUMAN STANDARD; PRT; 535 AA.  
ID Z257\_HUMAN AC Q9Y2Q1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).  
GN ZNF257 OR BMZF4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]\_SEQUENCE FROM N.A.  
RP TISSUE=Bone marrow;  
RC MEDLINE=20054457; PubMed=10585455;  
RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He X.-L.,  
RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;  
RT "Molecular cloning of six novel Kruppel-like zinc finger genes from  
RT hematopoietic cells and identification of a novel transregulatory  
RT domain KRBN";  
RL J. Biol. Chem. 274:35741-35748 (1999).  
CC -!- FUNCTION: May function as a transcription factor.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: Contains 1 KRAB domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF070651; AAD20957.1; -.  
DR HSP; P08047; ISP2.  
DR GENE; HGNC:13498; ZNF257.  
DR MIM; 606957; -.  
DR GO; GO:0005634; C:nucleus; NAS.  
DR GO; GO:0003677; F:DNA binding; NAS.  
DR GO; GO:0008270; F:zinc ion binding; NAS.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
DR InterPro; IPR001903; KRAB.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF01352; KRAB; 1.  
DR Pfam; PF00096; zf-C2H2; 10.  
DR PRINTS; PR00048; ZINC\_FINGER.  
DR ProDom; PD000003; Znf\_C2H2; 9.  
DR SMART; SM00349; KRAB; 1.  
DR SMART; SM00355; Znf\_C2H2; 12.  
DR PROSITE; PS00805; KRAB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 10.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 12.  
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;  
KW Nuclear protein; Repeat.  
FT DOMAIN 4 75 KRAB.  
FT ZN\_FING 173 195 C2H2-TYPE.  
FT ZN\_FING 201 223 C2H2-TYPE.  
FT ZN\_FING 229 251 C2H2-TYPE.  
FT ZN\_FING 257 282 C2H2-TYPE (DEGENERATE).  
FT ZN\_FING 288 310 C2H2-TYPE.  
FT ZN\_FING 316 338 C2H2-TYPE.  
FT ZN\_FING 344 366 C2H2-TYPE.  
FT ZN\_FING 372 397 C2H2-TYPE (ATYPICAL).  
FT ZN\_FING 403 425 C2H2-TYPE.  
FT ZN\_FING 431 453 C2H2-TYPE.  
FT ZN\_FING 459 481 C2H2-TYPE.  
FT ZN\_FING 487 509 C2H2-TYPE.  
SQ SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;

Query Match 16.8%; Score 374; DB 1; Length 535;  
Best Local Similarity 33.5%; Pred.No.3.le-15;  
Matches 88; Conservative 38; Mismatches 115; Indels 22; Gaps 9;  
Qy 8 KSISSLSSSSSRPK-----KYICTYEGCDKAYNRPSLLEOHLRTHSNDPRYKCTVDDC 62  
Db 149 KYVKFYKFSNDRHKIRHTTEKTKCKEKGKAFNQSSALTRKMTHTGKPKC--EBC 206  
Qy 63 DKAFPRKSHLETHIVSHSEKPPHCSVCGKGVNSROHLKHEITHTKSFCTPENQCEAP 122  
Db 207 GKAFNRSSHLTQHKVIHTREKPKYCEGKAFNRSSHTQHKRIHNRKPKYDECCKAF 266  
Qy 123 YKHOSL-----RHHLSVHEKTLCKQCNKVFTSPKLAQHLKHGSGPAYQCDHPGCFX 178  
Db 267 KWSALTTLTQHKRIHTGKPKYCEGKAFNQSSALTRKMTHTGKPKC--FQCEB--CGK 323  
Qy 179 NFQVSWVLQPHIKQSHPK---LKCPKCGKGVCKGLSSHMLSHDDSTMIKWTCD-YCD 234  
Db 324 AFNRSSHLTQHKVIHTREKPKYCEGKAFNRSSHTQHKRIH---TREKAYKDEYCK 379  
Qy 235 VGKFAKKNELVEHYNIFHDGNIP 257  
Db 380 AFNWSALTTLTQHKRIHTGKPK 402

Search completed: May 5, 2004, 15:21:55  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:19:16 ; Search time 20 seconds  
(without alignments)  
1981.545 Million cell updates/sec

Title: US-09-831-804-3

Perfect score: 2229  
Sequence: 1 MSSEDTKSISSLISSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	837.5	37.6	429	2 S20050	transcription fact
2	495	22.2	374	2 T37676	zinc-finger protei
3	400	17.9	803	2 S26823	zinc finger protei
4	396.5	17.8	595	2 G2075	transcription repr
5	395.5	17.7	363	2 I38937	DNA/RNA-binding pr
6	393.5	17.7	1191	2 S35305	zinc finger protei
7	375.5	16.8	1350	2 S00647	finger protein - A
8	374.5	16.8	710	2 I48668	zinc finger protei
9	373	16.7	542	2 A54651	zinc finger protei
10	373	16.7	651	2 B32891	finger protein 2,
11	372	16.7	393	2 JN0533	finger protein pML
12	364.5	16.4	338	2 G01496	transcription fact
13	364	16.3	428	2 A32891	finger protein 1,
14	363.5	16.3	615	2 S06546	finger protein (cl
15	359.5	16.1	427	2 A35659	krueppel-related p
16	354	15.9	693	2 I37570	zinc finger protei
17	353.5	15.9	591	2 S05088	finger protein XFO
18	351	15.7	474	2 I54338	zinc finger protei
19	348.5	15.6	439	2 S06556	finger protein (cl
20	348.5	15.6	707	2 S68858	finger protein - m
21	345.5	15.5	335	2 J01441	transcription fact
22	345.5	15.5	594	2 T12488	hypothetical prote
23	344	15.4	337	2 S05020	finger protein ZNF
24	344	15.4	580	2 A37107	spermatogenesis pr
25	343	15.4	686	2 A34612	zinc finger protei
26	342	15.3	488	2 S47072	finger protein HZF
27	341	15.3	336	2 S06578	finger protein (cl
28	341	15.3	347	2 S00549	developmental cont
29	340	15.3	339	2 J01442	transcription fact

30 339 15.2 411 2 S10245 finger protein, te  
31 338 15.2 247 2 S47070 finger protein HZF  
32 338 15.2 383 2 C92891 finger protein 9,  
33 337.5 15.1 728 2 A48830 probable transcrip  
34 337 15.1 469 2 I38600 zinc finger protei  
35 337 15.1 595 2 J07779 Krueppel-associate  
36 336 15.1 553 2 S22954 finger protein zfp  
37 335.5 15.1 428 2 S03677 finger protein (cl  
38 334.5 15.0 280 2 S06572 finger protein (cl  
39 332 14.9 536 2 S06548 finger protein (cl  
40 332 14.9 701 2 T14757 hypothetical prote  
41 331 14.8 794 2 S59069 Z13 protein - mous  
42 330.5 14.8 732 2 S47073 finger protein HZF  
43 329.5 14.8 496 2 T08674 probable finger pr  
44 329 14.8 589 2 I38598 zinc finger protei  
45 329 14.8 728 2 A54603 transcription fact

#### ALIGNMENTS

##### RESULT 1

S20050  
transcription factor IIIA - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein P9677.9; protein YPR186C; TFC2 protein  
C:Species: Saccharomyces cerevisiae  
C>Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text\_change 01-Dec-2000  
C:Accession: S20050; A44086; S58816  
R:Archambault, J.; Milne, C.A.; Schappert, K.T.; Baum, B.; Friesen, J.D.; Segall, J.  
J. Biol. Chem. 267, 3282-3288, 1992  
A:Title: The deduced sequence of the transcription factor TFIIA from Saccharomyces cere  
A:Reference number: S20050; MUID:92147684; PMID:1737784  
A:Accession: S20050  
A:Molecule type: DNA  
A:Residues: 1-429 <ARC>  
A:Cross-references: EMBL:M80611; NID:G172902; PIDN:AAB08014.1; PID:G172903  
R:Woychik, N.A.; Young, R.A.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3999-4003, 1992  
A:Title: Genes encoding transcription factor IIIA and the RNA polymerase common subunit I  
A:Reference number: A44086; MUID:92237295; PMID:1570325  
A:Accession: A44086  
A:Molecule type: DNA  
A:Residues: 1-429 <WOY>  
A:Cross-references: GB:M90638  
R:Miller, N.  
submitted to the EMBL Data Library, April 1995  
A:Description: The sequence of S. cerevisiae cosmid 9677.  
A:Reference number: S58816  
A:Accession: S58816  
A:Molecule type: DNA  
A:Residues: 1-429 <MIL>  
A:Cross-references: EMBL:U25841; NID:G786295; PID:G786305; MIPS:YPR186C  
C:Genetics:

A:Gene: SGD:EFZ1; TFIIA; TFC2  
A:Cross-references: SGD:S0006390; MIPS:YPR186C  
A:Map position: 16R  
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology  
C:Keywords: DNA binding; nucleus; transcription factor; zinc finger

Query Match 37.6%; Score 837.5; DB 2; Length 429;  
Best Local Similarity 43.7%; Pred No. 1.2e-43;  
Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;

QY 1 MSSEDTKSISSLISSSSSRPKYICTYEGCDKAYNRPSLLSQH-LRTHSNDRPKYC 57  
23 IGRSESSSELSLSTSRSSSSNRPKYTFCDYDGCDAKTRPSILTEHQLSVQGLRAFCQ 82  
QY 58 TVDDCCDAKFRKSHLETHIVSHSEKPKPHCSVCGKGVNSRQHLKHEITHTKSFCKTPE 117  
Db 83 --DKAKSVKSKSHLERHLYHSDTKPQCSYCGKGVTRQQLKHEVTHTKSFICPEEG 140  
QY 118 COEAFYKHQSLRHILSVHEKLTCKQCNKVFTRPSKLAQHKLKHGG--SPAYQCDHPG 175

Db 141 CNLRFYKHPQLRAHLISVHLHLKLTCHCNKSFQRPYRLNHNLSKHHDEVENPYOCTFAG 200

Qy 176 CFKNFOTWSVLOPHIKQSHPKLCPKCGKGVGKGLSHMLSHDDSTMWIKIWTCDYCDV 235

Db 201 CCKEFRIMQLOSHIKNDHPKLCPCSKPCVCGNGLOHMHIIHDSLVTKWKKCHICPD 260

Qy 236 GKFAKXNVELVHNIFH-DGNTDPDL-LKETEVEKLENLLDOGSKLNNLHELETKLVE 293

Db 261 MGFSRKHDLITGYSHTBEDIPLELKYISDIQQL--VODHGVQLGN-----SKHNE 312

Qy 294 EDEEDSDLEKRSVDRSDVSQAQSIKSFASLE-GSKSVSKLISNGKINCPKNKC 352

Db 313 QDEEKISNLRKXKRLTENNVEFLQNEVDLEKRLSENGENLNLNTVGRKYRCFYNNC 372

Qy 353 DMFMSREYDLRHL-----KWHDDNQLQRIEFLNSIKER 387

Db 373 SRTFTKRYEKYKHIDKHVKHELKILQ-----EKKE 404

RESULT 2

zinc-finger protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 01-Dec-2000

C/Accession: T37676

R:McDougall, R.C.; Rajandream, M.A.; Batteil, B.G.; Davis, P.; Churcher, C.M. submitted to the EMBL Data Library, October 1999

A:Reference number: Z21736

A:Accession: T37676

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-374 <MCD>

A:Cross-references: EMBL:AL132675; PIDN:CAB59689.1; GSPDB:GN00066; SPDB:SPAC144.09c

A:Experimental source: strain 972h; cosmid c144

C:Genetics:

A:Gene: SPDB:SPAC144.09c

A:Map position: 1

A:Introns: 50/3

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 22.2%; Score 495; DB 2; Length 374;

Best Local Similarity 32.8%; Pred. No. 5.7e-23;

Matches 123; Conservative 58; Mismatches 160; Indels 34; Gaps 13;

Qy 15 SSSSSRRPKYICTYEGCDKAYNRPSSLLEQHLRTHSNDRPYKCTVDDCDKAFRRKSHLET 74

Db 13 SKNLRSAKKIFHCPEECGKYSRPSSLLEQHLRTHSNDRPYKCTVDDCDKAFRRKSHLET 72

Qy 75 HIVSHSEKPFHCSV--CGKGVNSRQHLRKHETH--TKSFCKTFCNCEAFYKHQSLRH 130

Db 73 HRCNTNVPFSCDYDCAQPYTQQLERHIEVHRKPKPYACTWEGDCDFSKHQQLRS 132

Qy 131 HLISVHEKTL---TCKQCNKVTRPSKLAQHLKHGGSPAYQCDHPC--FKNFQTS 184

Db 133 HISACHTLLPYCTVQDCLELPATKQKLNHVNRAHEKIIISYSCFHESCVGHEGFKWS 192

Qy 185 VLQFHFKQSHPKLCPKCGKGVGKGLSHMLSHDDSTMWIKIWTCDYCDV----GKFAK 240

Db 193 QLQNHIREAHVP--SCSICGRQFKAHAHLRHVVHLH--QTLBERKTYHCPMEGCKKSFTR 249

Qy 241 KNELVHYNIFHDGNIP---DDLKETEYK-KLENLLDOGSKLNNLHELETKLVEDE 296

Db 250 SSALKGHSVITHEGNMAFCHDCSGTFGKYHMLORHLERGTCCKAHPYINECQIKHDG 308

Qy 297 EDEEDSDLEKRSVDRSDVS--AQRSIKSFASLEGSKSVSKLISNGKINCPKNCDR 354

Db 309 IEGVAIHQKEKELSNLSDVAKIINEVTG--HGK-----EAREYSCSPFECNY 358

Qy 355 MFSREYDLRHLKWH 369

Db 359 RFKRLYDMRHLNSH 373

RESULT 3

S26823

zinc finger protein ZNF43 - human

N:Alternate names: zinc finger protein kox27

C:Species: Homo sapiens (man)

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000

C/Accession: S26823; I37967; S10416

R:Lowering, R.; Trowdale, J.

Nucleic Acids Res. 19, 2921-2928, 1991

A:Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell li:

A:Reference number: S26823; MUID:91279444; PMID:1711675

A:Accession: S26823

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-803 <LOV>

A:Cross-references: EMBL:X59244; NID:938031; PIDN:CAA41932.1; PID:938032

R:Thiesen, H.J.

New Biol. 2, 363-374, 1990

A:Title: Multiple genes encoding zinc finger domains are expressed in human T cells.

A:Reference number: I37949; MUID:91145339; PMID:2288909

A:Accession: I37967

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 476-531 <THI>

A:Cross-references: EMBL:X52358; NID:934160; PIDN:CAA36584.1; PID:9330090

C:Genetics:

A:Gene: GDB:ZNF43; HTP6

A:Cross-references: GDB:128653

A:Map position: 19p13.1-19p12

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

C:Keywords: DNA binding; zinc finger

Query Match 17.9%; Score 400; DB 2; Length 803;

Best Local Similarity 31.9%; Pred. No. 7.5e-17;

Matches 121; Conservative 53; Mismatches 171; Indels 34; Gaps 19;

Qy 2 SESEDTKSISSLISSSSS--SRPKYICTYEGCDKAYNRPSSLLEQHLRTHSNDRPYKCTV 59

Db 367 TECGEAFSRSSNLTKHKIHTKPKYK--EECGKAFKWSKLTTEHKLTHTGKPKYC-- 422

Qy 60 DCDKAFPRKSHLETHIVSHSEKPKHSCVCGKGVNSRQHLRKHETH--KSFCKTFCN 117

Db 423 EECGKAFNFWSTTKNRIHTGKPKYKVCVCGKAFQFNSLTTHKRHTAEKPKYC--EE 480

Qy 118 COEAFYKHQSL-RHHLSVHEKTLTCKQCNKVTRPSKLAQHLKHGGSPAYQCDHPC 176

Db 481 CGAFSRSSNLTKHKIHTKPKYKBECCGKAFKWSKLTTEHKLTHTGKPKYC--C 537

Qy 177 FKNFQTSVLQFH--IKQSHPKLCPKCGKGVGKGLSHMLSHDDSTMWIKIWTCDYCD 234

Db 538 GKAFNHFSLTKHKRIHTGKPKYKBECCGKAFQTSNLTTHKKIHTH--TGKPKYKBECC 593

Qy 235 VGK-FAKNELVEHYNIFHDGNIPDDLKETEYKLENLLDOGSKLNNLHELETKLVE 293

Db 594 -GRAFTQSSNLTHKKI-HTGGKP--YKCECGKAFNQPSTLTTHKKIHTH--TGKPKYKBECC 647

Qy 294 E-DEEDEDSDLEKRSVDRSDS--MSAQRSIKSFASLEGSKSVSKLISNGKINCPKN 350

Db 648 ECGKAFKWSSTLTTHKKIHTGKPKYKBECCGKAF--KLSSTLTTHKKIHTGKPKYKBECC 704

Qy 351 NCDMFREYDLRHLKWH 369

Db 705 -CGKAFNRPNSNLIEHKIHTH 722

RESULT 4

G02075

transcription repressor zinc finger protein 85 - human

C:Species: Homo sapiens (man)

C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 01-Dec-2000

C/Accession: G02075

R:Poncellet, D.A.

submitted to the EMBL Data Library, September 1995

A;Reference number: G09169  
A;Accession: G02075  
A;Status: Preliminary; translated from GB/EMBL/DBD  
A;Molecule type: mRNA  
A;Residues: 1-595 <PON>  
A;Cross-references: EMBL:U35376; NID:g1017721; PIDN:AAA79179.1; PID:g1017722  
C;Genetics:  
A;Gene: GDB:ZNF85  
A;Cross-references: GDB:l132279  
A;Map position: 19p12-19p12  
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 17.8%; Score 396.5; DB 2; Length 595;  
Best Local Similarity 31.7%; Pred. No. 8.8e-17;  
Matches 120; Conservative 48; Mismatches 160; Indels 51; Gaps 17;

QY 10 ISSLISSSS--SRPKYICTYEGCDKAYNRPSLLEQHLRTHSDRPNKYCTVDDCDKAFPR 68  
DB 186 ISCTEHSRIHTRVNFYK--EECGKAFNWSSTLTKEHRIHTEGKPYKC--EECGKAFNQ 241  
QY 69 KSHLETHIVSHSEKKPHSCYCGGVNSRQHLKHEITHT--KSFKCTFENCQEAFFYKHQ 126  
DB 242 SSNLIKHKHIIHTEGKPYKCECGKAFNRFSLTTHKIIHTEGKPYKC--KECGKAFNRSS 299  
QY 127 SL-RHHILSVHEKILTKQCNKVTFRSKLAQHLKHGGSPAYQCDHPGCFNKFQTSV 185  
DB 300 TLTTHRKIIHTEGKPYKCECGKAFKQSSNLTTHKIIHTEGKPYKCKK--CGKAFNOSAH 356  
QY 186 LOFH--IKQSHPLKCPKCGKGVGKGLSHMLSHDDSTMKIWTCDYCDVGVKFAKNE 243  
DB 357 LTTHVEIHTGKPYKCEKCGKAFNHFSLTTHKIIH--TCEKPYKCEK--GKAFKHS 411  
QY 244 LVEHNYIPHDGNIIPDDLKETEYVKKLENLDQSKLNLHLETEKLVKVEDDEDESL 303  
DB 412 TLTTHKIIHTEGKPYKCECGKAFNQSGL-----TEHKKIHTGKPYKPYE-- 455  
QY 304 DEKSDVSDMSAQRSTKSTASL-----EGSK-----SVSKLSNSGKKINCPKN 350  
DB 456 CEKCKGKAFNQSNTTRHKHSITEKPYKCECGKGFKWPSTLTTHKIIHTEGKPYK--E 513  
QY 351 NCDRMFSREYDLRRHLKWH 369  
DB 514 ECGKAFNQSGLTTHKIIH 532

RESULT 5  
I38937  
DNA/RNA-binding protein - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
C;Accession: I38937  
R;Draw, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.  
Gene 159, 215-218, 1995  
A;Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIF.  
A;Reference number: I38937; MUID:95347600; PMID:7622052  
A;Accession: I38937  
A;Status: Preliminary; translated from GB/EMBL/DBD  
A;Molecule type: mRNA  
A;Residues: 1-363 <RES>  
A;Cross-references: EMBL:U20272; NID:g644870; PIDN:AAA75623.1; PID:g644871  
C;Superfamily: transcription factor IIF

Query Match 17.7%; Score 395.5; DB 2; Length 363;  
Best Local Similarity 31.1%; Pred. No. 5.9e-17;  
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

QY 8 KSISSL-----ISSSSSR-----PKYICTYEGCDKAYNRPSLLEQHLRTHSDR 53  
DB 7 ESVSLTIADAFIAGESSAPTPRPALPRFRFCISFPDSCSANYSKAWKLDHLCCKHTGR 66  
QY 54 PKYCTVDDCDKAFKSHLETHIVSHSEKSPHCSV--CGGVNSRQHLKHEH-EIHT-- 107  
DB 67 PFVDEYCGKAFIRDYHLRSLHTHTGKPFVCAANGCDQKFNKSLKXKHFKEKHQ 126

QY 108 TKSFKCTFENCQEAFFYKHQSLR-RHHILSVHEKILTKQ--CNKVFTRPSKLAQHLKHHG 164  
DB 127 QKQYICSDCKKTKTKKQQLKHOCQHTNEPLFKCTGCGGKGFASPSKLRHAKAHG 186  
QY 165 GSPAYQCDHPGCFNKFQTSVLFQHIKOSHPK----- 196  
DB 187 ----YVC-QKGSFVAKTTELLKGVRETHKBEILCEVCRKTFKKQKYLKQMKHTHAPER 241  
QY 197 --LKCPK--CGKGVGKGLSHMLSL-HDDSTMKIWTCDYCDVGVK-FAKQNELVEHNYI 250  
DB 242 DVCRCPRGCGTYTTFVNLQSHILSFHEES--RPFVCEHAGCGKTFAMKQSLTRH-AV 297  
QY 251 FHDGNIPDDLKETEYVK 269  
DB 298 VHD---PDKCKMKLVKK 312

RESULT 6  
S33305  
zinc finger protein ZNF91 - human  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence\_revision 02-Jun-1994 #text\_change 05-Nov-1999  
C;Accession: S33305  
R;Bellefroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; Ponce  
EMBO J. 12 1363-1374 1993  
A;Title: Clustered organization of homologous KRAA zinc-finger genes with enhanced expres  
A;Reference number: S33305; MUID:93223677; PMID:8467795  
A;Accession: S33305  
A;Molecule type: mRNA  
A;Residues: 1-1191 <BEL>  
A;Cross-references: EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186774  
A;Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residue 1  
C;Genetics:  
A;Gene: GDB:ZNF91; HPF7; HTF10  
A;Cross-references: GDB:132284  
A;Map position: 19p12-19p12  
C;Keywords: DNA binding; zinc finger

Query Match 17.7%; Score 393.5; DB 2; Length 1191;  
Best Local Similarity 35.8%; Pred. No. 2.8e-16;  
Matches 98; Conservative 40; Mismatches 109; Indels 27; Gaps 12;

QY 20 SRPKYICTYEGCDKAYNRPSLLEQHLRTHSDRPNKYCTVDDCDKAFKSHLETHIVSH 79  
DB 905 TREKPYKC--EECGKAFSQPSHLTTHKMHTEKPYKC--EECGKAFSQSSTLTTHKIIH 960  
QY 80 SEKKPFHSCVCGGVNSRQHLKHEITHT--KSFKCTFENCQEAFFYKHQSL-RHHILSVH 136  
DB 961 TGEKPYKCECGKAFKSKSTLTTHKIIHTEGKPYKC--EECGKAFSQSSTLTTHKIIH 1018  
QY 137 EKTLTCKQCNKVTFRPSKLAQHLKHGGSPAYQCDHPGCFNKFQTSVLFQHIKOSHPK 196  
DB 1019 EKPYPKECGKAFNRSSKLTTHKIIHTEGKPYKCEE--CGKAFISSSTLNGH-KRIHTR 1074  
QY 197 ---LKCPKCGKGVGKGLSHMLSHDDSTMKIWTCDYCDVGVKFAKQNELVEHNYI 253  
DB 1075 EKPYPKECGKAFSQSSTLTTHKIIH--TGEKPYKCEG--GKAFKSSSALTGKIIHT 1129  
QY 254 GNIPDDLKETEYVKLENLDQSKLNLHLEET 287  
DB 1130 GE-----RYPKCEKCKKAFNQSSTLTTHKIIH 1157

RESULT 7  
S00647  
finger protein - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 31-Dec-1993  
C;Accession: S00647  
R;Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.  
EMBO J. 6 3065-3070, 1987  
A;Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.







Query Match	16.1%; Score 359.5; DB 2; Length 427;
Best Local Similarity	32.0%; Pred. No. 1.1e-14;
Matches	99; Conservative 36; Mismatches 121; Indels 53; Gaps 13;
QY	20 SRPKYICTVEGCDKAYNRSLLLEOHLRTHSNDPRKYCTVDDCKAFPRKSHLETHIVSH 79
DB	132 TRVNFYKC--EAYGRANFWNSTLNKXKRLHTGEKPYKC--KECGKAFNQTHSLTRHKRIH 187
QY	80 SEKKPFPHCSVCGKGVNSRQHLKRHEITHTKS--FKCTPENCQEAIFYKHQS--RHHLISVH 136
DB	188 TEEXPYKCEECGKAFNQSSLTTHNIHTGEIPYKC--EKCVRAFNOASKLTEHKLHTG 245
QY	137 EKTU.TCQCNKVFRPRSKLAQHLKH-----HGGSPAY 169
DB	246 EKRYECEECGKAFNRSSKLTETHYIHTGEKLYKCECGKAFNQSSLTTHKRIIHSGRKPY 305
QY	170 OGDHGGCFKFNQTSVLIQF--IKOSHPLKCPKCGKGVCKGGLSSHMLSHDDSTMIKI 227
DB	306 KCEB--CGKAFKQNSLTDHKKLHTGEKPYKCECGKAFNQSLNTRHKVIH---TGEKP 360
QY	228 WTCDCYDVGKFAKKNELVEHYNIFPHGNIPTDDLKETEYVKLENLLDQSGKLNHLHELET 287
DB	361 YKCGEC--GKAFNQSSALNTHKIIHTGPNBHKRESGKVFL-----SSKLSCTCKKIHT 412
QY	288 -EKL-KVEE 294
DB	413 GEKLYKCEE 421

Search completed: May 5, 2004, 15:23:24  
Job time : 21 secs